DIFFERENTIAL REPLICATION LEADS TO COPY NUMBER VARIATION AND GENOME REARRANGEMENTS OF POLYPLOID CELLS

By

William Henry Yarosh

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, MD

October, 2014

Abstract

Although polyploidy is pervasive throughout nature, its function is still debated. It can arise through multiple mechanisms; one mechanism is the endocycle, a modified cell cycle where the genome is replicated, but mitosis does not occur. An endocycle does not require that the genome replicate completely. This can result in a phenomena known as under replication and results in copy number variation of the genome. In addition to copy number variation, it results in different genomic structure; however this structure is not known and there are multiple models describing the potential structures arising from under replication. One model assumes that the free ended DNA resulting from under replication are not resolved, resulting in a DNA net. Another model assumes that free ended DNA are resolved resulting in genomic rearrangements as the free ends are connected to each other, resulting in parallel strands of DNA of different sequences and length. This study uses paired end reverse terminator sequencing to measure copy number variation as well as identify genomic rearrangements in order to determine which regions of the genome under replicate and the structure of an under replicated genome in multiple polyploid cell types. The under replicating regions of the genome contain similar

types of chromatin, genes with similar function, and random rearrangements within these regions to resolve free ends of the DNA resulting from incomplete DNA replication. These findings provide insights into the function of copy number variation, and answer the question of the structure of the genome resulting from under replication. In addition, they suggest that genomic rearrangements could result in additional gene diversity or reprogram the expression of existing genes.

My advisor is Dr. Allan Spradling and the secondary reader and thesis committee member is Dr. Xin Chen. Dr. Samer Hattar and Dr. Joe Gall are additional committee members.

Acknowledgements

I would like to thank Dr. Allan Spradling for being an excellent mentor during my graduate education at Johns Hopkins University. I would like to thank Dr. Xin Chen, Dr. Joe Gall, Dr. Nicholas Ingolia, and Dr. Samer Hattar for being on my thesis committee and their guidance helping me overcome the challenges of my research. I would like to thank Allison Pinder for her expertise in deep sequencing. I would like to thank the members of the Spradling laboratory, past and present, for their support. I would like to thank Joan Miller for helping me navigate rules of multiple departments within the university.

Table of Contents

Chapter 1: Polyploidy and Differential Replication

Author Contribution

This chapter is solely my own work. I have written this chapter to provide background information on aspects of biology related to my thesis research and provide some context as to the status of the field prior to this work.

Cell Cycle

The cell cycle – nearly identical DNA replication followed by equal segregation of homologous chromosomes between daughter cells – occurs in distinct stages, and is regulated by several cell cycle checkpoints. These occur at the end of G1 before S, the end of G2 before M, and during mitosis during metaphase.

In order to bypass the G1-S checkpoint, cyclin D needs to interact with cyclin dependent kinase 4/6, which is competitively bound by cyclin dependent kinase inhibitor p14ARF; once together cyclin dependendent kinase 4/6 and cyclin D phosphorylate retinoblastoma protein which blocks it from inhibiting E2F which then allows expression of cyclin E which then interacts with cyclin dependent kinase 2 allowing for S phase to initiate (Bertoli et al. 2013). Although there are many factors regulating the key molecules governing this checkpoint, if cyclin E is expressed and can interact with cyclin dependent kinase 2 then the cell will pass this restriction point and enter S phase. During S phase euchromatic sequences are the first to initiate replication followed by heterochromatic sequences; this is the result of both the position of origins on the genome, the time at which they fire during S

phase, and the decreased rate of replication through some heterochromatic sequences (Edgar et al. 2001).

Before the genome proceeds into mitosis, the second cell cycle checkpoint must be bypassed. This checkpoint is normally blocked by signals indicating unfavorable conditions for mitosis, such as DNA damage and incomplete genome replication. If suitable conditions are present, cell division cycle 25 dephosphorylates the mitosis-promoting factor, which consists of cyclin B and cyclin dependent kinase 1. This in turn drives the cell cycle into M phase. If unsuitable conditions are present, these proteins can be phosphorylated blocking their activity. In order for M phase to occur, cyclin B and cyclin dependent kinase 1 need to be activated and processes inhibiting their activation will block mitosis.

During mitosis there is one final additional checkpoint, the metaphase checkpoint, where the cell verifies that the chromosomes have lined up properly and are under the proper tension before anaphase occurs. The anaphase promoting complex is activated if these conditions are met and cyclin B is degraded as well as securin which inhibits seperase which breaks down the cohesisns holding the sister chromatids together (Ciosk et al. 1998). This complex can also down regulate other cyclins.

Genome Repair

The cell is able to detect many different types of DNA damage. During DNA replication, polymerases can incorporate the incorrect bases into the sequence, and DNA polymerase, can repair some of these errors. DNA polymerase can also fail to replicate part of the genome due to replication fork stalling and collapse. This will produce free-ended DNA that will be detected by the cell, and the cell will try to repair them through either homologous recombination or microhomology-mediated end joining if it is possible, but it may need to reply on non-homologous end joining. Non-homologous end joining can lead to ligation of the incorrect free ends causing rearrangement and potentially loss of DNA sequence. Non-homologous end joining is often accompanied by variant histone 2a which binds to regions thought to contain stalled or collapsed forks and then directs the cell to repair the free ends at those locations via ligation (Andreyeva et al 2008).

Endocycles

Nuclear DNA content is not a completely stable constant in all cells. A change in the course of the cell cycle can result in polyploidy if the nuclear DNA is

replicated, but daughter cells are not formed. Progression through a sequence of G→S phases, endocycles, increases the nuclear DNA content within a cell and can also produce polytene chromosomes that consist of multiple chromosomes associated with each other (Beerman 1962); notwithstanding, some endocycling polyploid cells may not undergo chromosome polytenization even in the tissues containing polytene cells in the same individual (Henderson 1967). Once a cell has become polyploid it is not limited to utilizing just endocycles; polyploid cells can subsequently undergo cell divisions (Fox et al. 2010; D'Amato et al. 1948; Grell 1946) or modify endocycles so that they progress through G→S→aborted M phases (Poroshenko et al. 1970; Moorhead et al. 1956). Endocycles provide a mechanism to alter the nuclear DNA content thereby altering the cell; naturally, many tissues of a number of eukaryotic species utilize endocycles during development (Heitz 1933; Spradling et al. 1980a; Edgar et al. 2001).

Switching From the Cell Cycle to the Endocycle

In order to switch from the cell cycle to the endocycle, a cell needs to repress mitosis while still utilizing many of the proteins required for DNA replication related to the G1–S restriction point. (Edgar et al. 2001;

Lee et al. 2003). In order to bypass mitosis, the mitotic cyclins can be degraded by the highly conserved anaphase promoting complex; however, mitotic cyclins are also not highly expressed in endocycling cells (Sigrist et al. 1997, Lilly et al. 1996). The degradation of mitotic cyclins by the anaphase promoting complex inhibits cyclin dependent kinase 1 activity promoting the endocycle.

In order to trigger genome replication cyclin E levels rise and fall again. This produces an oscillation of cyclin E activity that can be achieved through multiple mechanisms, but is the driving force of the endocycle. For example, increasing cyclin E levels is accompanied by the accumulation of dacapo in the *Drosophila* ovary nurse cells (Bell et al. 2002). Eventually, dacapo levels rise high enough to inhibit Cyclin E and cyclin dependent kinase 2 activity stopping S phase and allowing the cell to reenter G phase due to low cyclin dependent kinase activity and allow origins to be relicensed for another round of replication during the next S phase (Bell et al. 2002). Another example, in breast cancer cell lines human F-box protein hCdc4 targets cyclin E for degradation, and there are other mechanisms by which cyclin E oscillation can be achieved (Strohmaier et al. 2001).

The combination of the inhibition of mitotic cyclins and the oscillation of cyclin E activity allow the cell to bypass mitosis and switch between S and G phase producing an endocycle.

Nuclear Polyploidy in Nature

Alteration of gene copy number in a cell is a common developmental strategy accompanying differentiation in specific tissues in many eukaryotic species; however, the degree of polyploidy and occurrence of polyteny is species, tissue, and cell specific. Polyploidy and polyteny occur frequently in seeded plants, in particular flowering plants (Anisimov 2005; D'amato 1964). The degree of polyploidy can vary among cells in the same tissues (Nagl 1974); in addition, ploidy variation can be present in striking patterns suggesting a role in development. The suspensor cells of *Phaseolus* show a gradient pattern in the ploidy of cells from high levels of polyploidy to diploid (Nagl 1974). Alteration of nuclear DNA content is not limited to multi-cellular eukaryotes as it occurs in the simplest animals. Some protists, such as Percolozoa, are polyploid, but consist of multiple diploid nuclei (Corliss et al. 1982).

Some of the most interesting polyploidy cells are found in ciliated protozoans. Ciliates have a polyploid macronucleus which develops from the diploid micronucleus following conjugation and meiosis (Nagl 1976; Nanney 1980; Gall 1984). The macronuclear precursor undergoes multiple endocycles that generate polytene chromosomes (Sonneborn 1977). The polytene chromosomes of ciliates display a light-dark banding pattern (Spear et al. 1976). The banding pattern becomes evident as a result of increasing alignment of corresponding chromatin segments and bundling of individual homologous chromosomes (Beerman 1962); the difference in contrast is the result of differences in physical density of chromatin (Engstrom et al. 1951). These polytene chromosomes then fragment into vesicles where some of the DNA is degraded (Spear et al. 1976). The remaining DNA is liberated from the vesicles and undergoes additional rounds of DNA synthesis – amplification of specific sequences – followed by splicing and rearrangement of newly synthesized DNA and remaining DNA of the original polytene chromosomes in a nuclear envelope to form the macronucleus (Nagl 1976). This results in differential representation of these polytene chromosomes, and since the DNA of the macronucleus is used to encode RNA it also directly affects gene expression (Nagl 1976).

There are also many examples of polyploidy, polyteny, and differential replication of nuclear DNA in multi-cellular animals. One example, megakaryocytes, in humans and other mammals differentiate in response to thrombopoietin; this developmental cue causes these cells to become polyploid before forming platelets (Nagata et al. 1997). Polytene cells have been observed in the trophoblast of some mammals (Varmuza et al. 1988), but the individual chromosome strands do not associate with each other tightly. In humans, the association of the individual chromosomes in these cells has only been seen transiently, during a short period of the endocycle following replication (Sarto et al. 1982). Polytene chromosomes in cells that undergo an aborted M phase, can disperse and form multiple individual chromosomes (Nagl 1970); these mitotic like changes account for the differences in appearance of chromosomes between the trophoblast cells of different mammalian species (Zybina et al. 1996).

The most striking and well-characterized examples of polytene chromosomes are those of the salivary glands of the *Drosophila* larva; they produce polytene chromosomes with a clear and consistent banding pattern (Urata et al. 1995).

Under-replication

Endocycling cells can also differentially replicate portions of their chromosomes through either amplification or by under replicating these regions during polyploidization; furthermore, differential replication can be restricted to specific chromosomal regions which manifest as a result of both general features of DNA replication (Laird 1973) and under tissue specific control (Nordman et al. 2010; Moshkin et al. 2001). During DNA replication euchromatic sequences are the first to initiate replication followed by heterochromatic sequences (Edgar et al. 2001). DNA synthesis during the S phase of the endocycle can result in the under-replication of certain DNA sequences. If this occurs heterochromatic sequences may not finish replication. Heterochromatin is under represented in polytene nuclei when compared to diploid nuclei (Heitz 1933). Heterochromatin contains highly repetitive DNA, such as satellite DNA, are under replicated (Heitz 1934; Gall et al. 1971; Vig et al. 1988). Heterochromatin also contains genes, pseudo-genes, arrays of ribosomal DNA and other non-coding RNA's flanked by DNA repeats, mobile DNA elements, and is thought to contain most unmapped DNA (Smith et al. 2007). These other components (e.g. ribosomal DNA) (Zuchowski et al. 1976) and

intercalary heterochromatin (e.g. Ultrabithorax) can be under-replicated in endocycling cells (Moshkin et al. 2001). Euchromatin might also be under-replicated due to the position effects of neighboring heterochromatin (Ananiev et al. 1974; Karpen et al. 1990). P-Element insertion of *rosy* sequences into the telomeric heterochromatin of *Drosophila* salivary gland polytene chromosomes is under-replicated along with the telomeres (Karpen et al. 1992). Interestingly, lower levels of Cyclin E activity allow heterochromatin, normally under-replicated to be replicated; it is possible that low levels of Cyclin E allow more time for DNA replication (Lilly et al. 1996). Other cyclins required in the mitotic cell cycle are not required in endocycles and are absent in endocycling cells (Edgar et al. 2001). Only Cyclin E (Lehner et al. 1990; Stem et al. 1993) - which can down regulate itself in the endocycle, but not during the mitotic cell cycle (Duronio et al. 1995; Sauer et al. 1995) — has been demonstrated to effect endocycling cells; this provides a mechanism for endocycles to persist through successive S->G phases (Lilly et al. 1996). Mutation of Suppressor of Under-Replication (SuUR), which binds to heterochromatin, inhibits under-replication in endocycling cells (Pinyurin et al. 2008).

Amplification

Amplification is usually reported in tissues which require high expression of specific genes, but cannot transcribe sufficient quantities of RNA from these genes since cells are limited by template, transcription rates, and developmental timing (Spradling et al. 1980a). In frog oocytes, which are exceptionally large, ribosomal cistrons are amplified (Gall et al. 1969; Barsacchi et al. 1972; Stahl et al. 1975; Raikova et al. 1979). Amplification of specific chromosome segments is not limited to germ cells, for example the follicle cells of the *Drosophila* ovary amplify several genes which are crucial for proper egg chamber development (Spradling et al. 1980b; Claycomb et al. 2004). In these cells eggshell (chorion) gene amplification occurs by repeated firing of replication origins within a single S phase. These origins are regulated by cis acting elements (De Cicco et al. 1984) and require normal initiation factors and replication proteins (e.g. ORC, MCM, and PCNA) (Bosco et al. 2002; Claycomb et al. 2004). The degree of amplification peaks at the centers of these replication clusters and decreases as the distance increases from the center of the cluster (~50 kb) (Spradling 1981). Chorion genes are located in two major clusters, s38 and 36 are on X chromosome at 7F and s19 and

15 on the 3rd chromosome at 66D which show ~15 fold and ~60 fold amplification respectively (Spradling 1981). Two other amplicons on the 2nd chromosome at 30B and on the 3rd at 62D were identified recently using euchromatic array CGH in follicle cells (Claycomb et al. 2004); furthermore, this study also demonstrated that genes in these amplicons are expressed and one of the genes, yellow g (located in 62D), is essential for eggshell formation (Claycomb et al. 2004).

Under Replication in the Drosophila Larval Salivary Glands

There are many examples of differential replication in insects, in particular Dipters, which utilize polyploidy and polyteny extensively in their tissues throughout development (Nagl 1976; Endow et al. 1975; Smith et al. 1991). One of the most well studied examples are the salivary glands of the *Drosophila* larva. They arise from an invagination of two ventral ectodermal primordial after about 7 hours of development and consist of 100 cells each (Haberman et al. 2003). Although the salivary glands consist of 3 cell types, salivary gland cells, ring cells, and duct cells, the salivary gland cells enter the endocycle and increase in genome copy number dramatically (Haberman et al. 2003). They go through 10 endocycles which are not synchronous and have many regions of under

replication, in particular the heterochromatin (Endow et al. 1975); however, some euchromatin is also under replicated (Nordman et al. 2010). Although this system has been studied extensively for decades, the number and location and depth of under replication is not known.

Differential replication in the Drosophila ovary

The *Drosophila* ovary contains both germline and somatic cells which exhibit differential replication – exhibiting both differential under-replication and amplification – and provides a model system to study polyploidy, polyteny and differential replication; furthermore, it has already contributed a great deal to our current understanding of characteristics, function, and regulation of endocycling in cells (Spradling 1993a; Dej et al. 1999; De Cicco et al. 1984). The *Drosophila* ovary consists of ovarioles containing both germline and somatic cells (Spradling 1993a). These two cell populations develop together as follicles (or egg chambers) over the course of 14 distinct stages (Spradling 1993a). These cells arise from two stem cell populations, the germ line stem cells (De Cuvas et al. 1997) which give rise to cystoblasts that undergo 4 mitotic divisions give rise to the oocyte interconnected with 15 nurse cells (Bohrmann et al. 1993),

and the follicle stem cells (Nystul et al. 2010) which also initially divide by mitosis to give rise to nearly a thousand follicle cells forming a monolayer over the germ cells. After mitosis the nurse cells switch to the endocycle and form polytene chromosomes as they completely reduplicate their DNA 5 times. Then, in the stage 5 egg chamber, they undergo an endocycle with an incomplete M phase each polytene chromosome to disperse into 32 chromosome pairs (Dej et al. 1999). The nurse cells then undergo 5 more endocycles with an incomplete S phase (Lilly et al. 1996) causing heterochromatic regions to be under-replicated (Dej et al. 1999). Nurse cells trigger nuclear envelope breakdown during stage 10B and transfer their contents to the developing oocyte (Royzman et al. 2002). The follicle cells, which have formed a monolayer around the germ cells, also switch to the endocycle – they undergo 3 endocycles during follicle stages 7-9 – and under-replicate their heterochromatin (Sun et al. 2005). At stage 10 the follicle cells stop endocycling and switch to an even more extreme form of chromosome sequence copy number change as they begin amplifying specific gene clusters, in particular those containing the chorion genes which will be in high metabolic demand during the final 10 hours of follicle development (Spradling et al. 1980a).

Structural consequences of differential replication

Differential replication of polytene chromosomes raises a problem since newly synthesized DNA is not continuous with the template (Nagl 1976). If replication forks are prematurely terminated then it is possible that a DNA net with multiple branch points will be formed by replication forks at different locations (Laird 1973). Amplified rDNA forms extra-chromosomal molecules (Hourcade 1973), but they could be integrated into the chromosome altering the sequence (Biedler et al. 1976; Sager et al. 1985; Tartof 1971). The DNA net model was proposed originally to explain the structure of under-replicated chromosomes (Laird 1973), but such structures have only been visualized in the case of chorion gene amplification (Osheim et al. 1983). Replication forks have not been detected in under-replicating polytene chromosomes (Nagl 1976; Sorsa 1976); furthermore, in the Dp1187 mini-chromosome, which exhibits a gradient of under-representation expected in the DNA net model, does not have replication forks (Karpen et al. 1992). This suggests that the under-replication of some heterochromatin is the result of partial DNA elimination followed by ligation of free double strand breaks (Karpen et al. 1992). The simplest

mechanism would be fork breakage and religation at each endocycle. Furthermore, DNA elimination and chromosome reorganization occurs during in the formation of the macronucleus of ciliates (Nagl 1976). Amplification or under-replication producing a DNA net could also produce free DNA molecules if the net is transient, releasing the newly synthesized single stranded DNA or incomplete replication formed double stranded DNA breaks at the replication forks that collapse (-2001). Free double stranded DNA would exist transiently and be removed or repaired and ligated to other DNA molecules, and propagated through DNA replication (Karpen et al. 1992).

The incredible changes in the structure of the polytene chromosome during *Drosophila* larval development suggest that there are changes to the genome sequence occurring under replication. The first would be the formation of the chromocenter, the region of the genome where all the chromosomes merge. Although the sequence of this region is not known, it is thought to be largely made of heterochromatin (Usov et al. 2011). Although it's possible that the chromosomes interact with each other through non-covalent interactions to form the chromocenter, it is also possible that this structure arises due to genomic rearrangements between under replicated

heterochromatin. Other evidence of structural changes are the observations of ectopic fibers found in a wide range of polytene chromosome preps that occur at the same regions. These could be inter chromosomal rearrangements of a few strands. They could also be free ended strands intermittently associating with regions of other chromosomes. Another interesting observation is the toroidal structure of a fraction of the bands of the polytene chromosome. It is interesting that these tend to associate with regions, which tend to be under replicated (Sorsa 1983). It's possible that the lack of replication alone is enough to account for these structural changes, but they could also arise from the super structure accommodating rearrangements between the free ends generated by under replication.

References

Ananiev E.V. and Gvozdev V.A. (1974): Changed pattern of transcription and replication in the polytene chromosomes of Drosophila melanogaster resulting from eu-heterochromatin rearrangement. Chromosoma. 45(2):173-91.

Andreyeva E.N., Kolesnikova T.D., Belyaeva E.S., Glaser R.L., Zhimulev I.F. 2008. Local DNA underreplication correlates with accumulation of phosphorylated H2Av in the Drosophila melanogaster polytene chromosomes. Chromosome Res. 16: 851-62.

Anisimov A.P. (2005): Endopolyploidy as a morphogenetic factor of development. Cell Biol Int. 29(12):993-1004.

Barsacchi G. and J.G. Gall. (1972): Chromosomal localization of repetitive DNA in the newt, Triturus. J. Cell Biol. 54(3): 580-91.

Beerman, W. (1962): Reisenchromosomen. Protoplasmatologia VI/D. Wein: Springer

Bell S.P. and Dutta A. (2002): DNA replication in Eukaryotic Cells. Annual Review of Biochemistry. 71: 333-374.

Bertoli C.B., Skotheim J.M., Robertus A.M. de Bruin. (2013). Control of cell cycle transcription during G1 and S phases. Nature Reviews Molecular Cell Biology. 14: 518-528.

Biedler J.A. and Spengler B.A. (1976). Metaphase chromosome anomaly: association of drug resistance and cell specific products. Science. 191: 185-187.

Belyaeva E.S., Zhimulev I.F., Volkova E.L., Alekseyenko A.A., Moshkin Y.M, Koryakov D.E. (1998): Su(UR)ES: A gene suppressing DNA underreplication in intercalary and pericentric heterochromatin of Drosophila melanogaster polytene chromosomes. P.N.A.S. 95(13): 7532-537.

Bennxci A., Buiatti M., D'Amato F., Pagliali M. (1971): Nuclear behavior in Haplopappus gracilis callus grown in vitro on different culture media in Les Cultures de Tissus de Plantes. Colloques Internationaux du C.N.R.S. 193: 245-250.

Bohrmann J and Haas-Assenbaum A. (1993): Gap junctions in ovarian follicles of Drosophila melanogaster: inhibition of promotion of dye-coupling between oocyte and follicle cells. Cell Tissue Research 273(1): 163-73).

Bosco G. and Orr-Weaver T.L. (2002): Regulation of the cell cycle during oogenesis and early embryogenesis in Drosophila. In: Regulation of Gene Expression at the Beginning of Animal Development. Ed. De Pamphilis M. Elsevier, Amsterdam: 107—154.

Ciosk, R., Zachariae, W., Michaelis, C., Shevchenko, A., Mann, M., Kim., N. (1998): An ESP1/PDS1 Complex Regulats Loss of Sister Chromatid Cohesion at the Metaphase to Anaphase Transition in Yeast. Cell 93(6): 1064-1076.

Claycomb J.M, Benasutti M, Bosco G., Fenger D.D., Orr-Weaver T.L. (2004): Gene amplification as a developmental strategy: isolation of two developmental amplicons in Drosophila. Dev. Cell 6(1): 145-155.

Corliss J.O. and Lipscomb D.L. (1982): Establishment of a new order in kingdom Protista for Stephanopogon, long-known

"ciliate" revealed now as a flagellate. Journal of Protozoology 92: 294.

D'amato F. and Avanzi M.G. (1948): Reazioni di natura auxinica ed effetti rizogeni in Allium Cepa L. Plant Biosystems 55(2) 161-213.

D'Amato F (1964): Endopolyploidy as a factor in plant tissue development. Caryologia 17: 41-52.

DeCicco D.V. and Spradling A.C. (1984): Localization of a cis-acting element responsible for the developmentally regulated amplification of Drosophila chorion genes. Cell 38(1):45-54.

De Cuvas M., Lilly M.A., Spradling A.C. (1997): Germline cyst formation in Drosophila. Annu. Rev. Genet. 31: 405–428

Dej K.J. and Spradling A.C. (1999): The endocycle controls nurse cell polytene chromosome structure during Drosophila oogenesis. Development. 126(2):293-303.

Duronio R.J. and O'Farrel P.H. (1995): Developmental control of the G1 to S transition in Drosophila: Cyclin E is a limiting downstream target of E2F. Genes & Dev. 9: 1456-1468.

Edgar B.A. and Orr-Weaver T.L. (2001): Endoreplication Cell Cycles: More for Less. Cell. 105(3): 197-306

Engstrom A. and Ruch F. (1951): Distribution of mass in salivary gland chromosomes. P.N.A.S. 37: 459-461.

Endow S.A. and Gall J.G. (1975): Differential replication of satellite DNA in polyploid tissues of Drosophila virils. Chromosoma. 50: 175-192.

Ferguson D.O., Sekiguchi J.M., Chang S., Frank K.M., Gao Y., DePinho R.A., Alt F.W. (2000): The nonhomologous end-joining pathway of DNA repair is required for genomic stability and the suppression of translocations. P.N.A.S. 97(12): 6630-6663.

Gall J.G., Macgregor H.C., Kidston M.E. (1969): Gene amplification in the oocytes of Dytiscid water beetles. Chromosoma. 26(2): 169-87.

Gall J..G. (1984): Molecular genetics: Ciliates come of age. Nature. 310: 453-454.

Glaser R.L, Leach T.L, Ostrowski S.E. (1997): The Structure of Heterochromatic DNA is Altered in Polyploid Cells of Drosophila melanogaster. Molecular and Cellular Biology. (13)3: 1254-1263.

Glover D. and Endow S.A. (1979): Differential replication of Ribosomal Gene repeats in polytene nuclei of Drosophila. Cell 17: 597-605.

Grell (1946): Cytological studies in Culex; somatic reductive divisions. Genetics. 31:60-76

Hammond, M. P. and Laird, C. (1985): Chromosome structure and DNA replication in nurse and follicle cells of Drosophila melanogaster. Chromosoma. 91: 267-278.

Heck M.M. and Spradling A.C. (1990): Multiple replication origins are used during Drosophila Chorion gene amplification. Journal of Cell Biology. 110(4):903-914.

Heitz E. (1933) Uber totale und partielle somatische Heteropyknose, sowie strukturelle Geschlechtschromosomen bei Drosophila funebriz (Cytologische Unterscuchungen an Dipteren, II). Anat. 18: 720-742.

Heitz E. (1934). Uber α and β- Heterochromatin Sowie Konstanz und Bau der Chromomeren bei Drosophila. Biol. Zentralbl. 54: 588-569.

Henderson S.A. (1967): The Salivary Gland Chromosomes of Dasyneura crataegi (Diptera: Cecidomyidae). Chromosoma. 23(1): 38-58.

Hourcade D., Dressler D., Wolfson J. (1973). The amplification of ribosomal RNA genes involves a rolling circle intermediate. P.N.A.S. 70: 2926-2930

Ingle J. and Timmis J.N. (1975): A role for differential replication of DNA in development. In: Modification fo the Information Content of Plant Cells Ed. Markham R, Davies D.R., Hopwood D.A. Horne R. W. Plant Physio. 55: 496-501.

Karpen G.H. and Spradling A.C. (1990): Reducing DNA polydenization of a minichromosome region undergoing

position-effect variegation in Drosophila. Cell. 63: 97-107.

Karpen G.H. and Spradling A.C. (1992): Analysis of subtelomeric heterochromatin in the Drosophila minichromosome Dp1187 by single P element insertional mutagenesis.

Kidder G.M. (1976): The Ribosomal RNA cistrons in clam gametes. Developmental Biology. 46(1): 132-142.

Kuzminov A. (2001): Single-stranded interruptions in replicating chromosomes causes double-stranded breaks. P.N.A.S. 98(15): 8241-8246.

Lane M.E., Sauer K., Wallace K., Jan Y.N., Lehner C.F., Vaessin H. (1996): Dacapo, a cyclin-dependent kinase inhibitor, stops cell proliferation during Drosophila development. Cell. 87(7): 1225-1235.

Laird C.D. (1973): DNA of Drosophila chromosomes. Annu. Rev. Genet. 7: 177-204.

Lamb M. and Laird C. (1987): Three euchromatic DNA sequences under-replicated in polytene chromosomes of Drosophila are localized in constrictions and ectopic fibers. Chromosoma. 95: 227-235.

Leach T.J., Chotkowski H.L., Wotring M.G, Dilwith R.L., and Glaser R.L. (2000): Replication of Heterochromatin and Structure of Polytene Chromosomes. Molecular and Cellular Biology. (20)17: 6308-6316.

Lee L.A. and Orr-Weaver T.L. (2003): Regulation of Cell Cycles in Drosophila Development: Intrinsic and Extrinsic Cues. Annual Review of Genetics. 37: 545-578.

Lehner C.F. and O'Farrel P.H. (1990): The roles of Drosophila cyclins A and B in mitotic control. Cell 61: 535-47.

Lilly M.A. and Spradling A.C. (1996): The Drosophila Endocycle is controlled by Cyclin E and lacks a checkpoint ensuring S-phase completion. Genes and Development. 10(19): 2514-2526.

Mardis E.R. (2008): Next-Generation DNA Sequencing Methods. Annual Review of Genomics and Human Genetics. 9: 387-403.

Moorhead, P.S. and Hsu T.C. (1956): Cytologic studies of HeLa, a strain of human cervical carcinoma. III. Durations and characteristics of the mitotic phases. J. Natl. Cancer Inst. 16(5):1047-66.

Moshkin Y.M., Alekseyenko A.A., Semeshin V.F., Spierer A., Spierer P., Makarevich G.F., Belyaeva E.S., Zhimulev I.F. (2001): The bithorax complex of Drosophila melanogasterL Underreplication and morphology in polytene chromosomes. P.N.A.S. 98(2):570-74.

Nagata, Y., Muro, Y., and Todokoro, K. (1997): Thrombopoietin- induced polyploidization of bone marrow megakaryocytes is due to a unique regulatory mechanism in late mitosis. J. Cell Biol. 139: 449–457.

Nagl W. (1970): Inhibition of polytene chromosome formation in Phaseolus by polyplid mitosis. Cytologia 35: 252-258.

Nagl W. (1974): The Phaseolus suspensor and ist polytene chromosomes. Z. Pflanzenphysiol. 73:1-44.

Nagl W. (1976): Endopolyploidy and Polyteny in Differentiation and Evolution. North Holland

Nanney D. L. (1980): Experimental Ciliatology an introduction to the genetic and developmental analysis in ciliates. New York: John Wiley and Songs.

Nordman J., Li S., Eng T., Macalpine D., Orr-Weaver T.L. (2010): Develomental control of the DNA replication and transcription programs. Genome Res. [Epub ahead of print]

Nystul T. and Spradling A. (2010): Regulation of epithelial stem cell replacement and follicle formation in the Drosophila ovary. Genetics. 184(2) 503-515.

Osheim Y.N., Miller O.L. Jr., Beyer A.L. (1988): Visualization of Drosophila melanogaster chorion genes undergoing amplification. Mol. Cell Biol. 8(7) 2811-21.

Pfitzer P. (1971): Polyploid nuclei in myocardial cells of the pig. Virchows Arch B Cell Pathol. 9(2):180-6.

Pinyurin A.V., Boldyreva L.V., Shloma V.V., Kolesnikova T.D., Pokholkovaa G.V., Andreyeva E.N., Kozhevnikova E.N., Ivanoschuk I.G., Zarutskaya E.A., Demakov S.A., Gorchakov A.A., Belyaeva E.S., Zhimulev I.F. (2008): Interaction between the Drosophila heterochromatin proteins SUUR and HP1. Journal of Cell Science. 121: 1693-1703.

Poroshenko G.G., Fomina M.M., Nikol'skaia T.A. (1970): Diplochromosomes in endoreduplication. Tsitologiia. 12(12):1575-8.

Prescott D.M. (1994): The DNA of ciliated protozoa. Microbiol. Mol. Miol. Rev. 58(2): 233-267.

Raikova E.V., Steinert G., Thomas C. (1979): Amplified ribosomal DNA in meiotic prophase oocytes nuclei of acipenserid fishes. Development of Genes and Evolution. 183(1): 81-85.

Ravid K., Lu J., Zimmet J.M., Jones M.R. (2002): Roads to polyploidy: The megakaryocyte example. Journal of Cellular Physiology. 190(1): 7-20

Royzman I., Hayashi-Hagihara A., Dej K.J. Bosco G., Lee J.Y., Orr-Weaver T.L. (2002): The E2F cell cycle regulator is required for Drosophila nurse cell DNA replication and apoptosis. Mechanisms of Development. 119(2): 225-237.

Sager R., Gadi I.K., Stephens L, Grabowy C.T. (1985): Gene amplification: an example of accelerated evolution in tumorigenic cells. P.N.A.S. 82(20): 7015-19

Sarto G.E., Stubblefield P.A., and Therman E. (1982): Endomitosis in the human Trophoblast. Human Genetics. 62(3): 229-232.

Sauer K., Knoblich J.A., Richardson H., Lehner C.F. (1995): Distinct modes of cycling E..cdc2c kinase regulation and S-phase control in mitotic and endoreduplication cycles of Drosophila embrogeneis. Genes & Dev. 9: 1327-339.

Sigrist J.S. and Lehner C.F. (1997): Drosophila fizzy-related Down-Regulats Mitotic Cyclins and Is Required for Cell Proliferation Arrest and Entry into Endocycles. Cell. 90(4): 671-681.

Smith A.V. and Orr-Weaver T.L. (1991): The regulation of the cell cycle during Drosophila embryogenesis: the transition to polyteny. Development. 112: 997-1008.

Smith C.D., Shu SQ., Mungall J.C., Karpen G.H. (2007): The Release 5.1 Annotation of Drosophila melanogaster Heterochromatin. Science. 316: 1586-1591.

Sonneborn T.M. (1977). Genetics of cellular differentiation: stable nuclear differentiation in eukaryotic unicells. Ann. Rev Genet. 11: 349-367.

Sorsa V. (1976): Beaded organization of chromatin in the salivary gland of Drosophila melanogaster. Hereditas. 84:213-220.

Sorsa V. (1983): Toroidal bands in polytene chromosomes of Drosophila. J. Cell Sci. 64:255-264.

Spear B.B. and Lauth M.R. (1976): Polytene chromosomes of Oxytricha: Biochemical and morphological changes during macronuclear development in a ciliated protozoan. Chromosoma. 54(1): 1-13.

Spradling A.C., Mahowald A.P. (1980a): Amplification of genes for chorion proteins during oogenesis in Drosophila melanogaster. P.N.A.S (77)2: 1096-1100

Spradling A.C., Digan M.E., Mahowald A.P. (1980b): Two clusters of genes for major chorion proteins of Drosophila melanogaster. Cell 19(4): 905-915.

Spradling A.C. (1981): The organization and amplification of two chromosomal domains containing Drosophila chorion genes. Cell 1(2):193-201.

Spradling A.C. (1993a): Developmental genetics of oogenesis, pp 1-70. In: The Development of Drosophila melanogaster. eds. Bate M. and Arias M. Cold Spring Harbor Press.

Spradling A.C. (1993b): Position Effect Variagation and Genomic Instability. Cold Spring Harbor Symposia on Quantitative Biology. LVIII: 585-595.

Stahl A., Luciani J.M., Devictor M., Capodano A.M., Gagné R. (1975): Constitutive heterochromatin and micronucleoi

in the human oocytes at the diplotene stage. Humangenetik. 26(4):315-27.

Stem B., Ried G., Clegg N.J., Grigliatti T.A., Lehner C.F. (1993): Genetic analysis of the Drosophila cdc 2 homolog. Development. 117: 219-232.

Storchova Z. and Pellman D. (2004): From polyploidy to aneuploidy, genome instability and cancer. Nat Rev Mol Cell Biol. 5(1): 45-54.

Sun J., Deng W.M. (2005): Notch-dependent downregulation of the homeodomain gene cut is required for the mitotic cycle/endocycle switch and cell differentiation in Drosophila follicle cells. Development 132: 4299-308.

Tautvydas K.J. (1976): Evidence for Chromosome Endoreduplication in Eudornia California, a Colonial Alga. Differentiation. 5(1):35-42.

Tartof K.D. (1971): Increasing the multiplicity of ribosomal RNA genes in Drosophila melanogaster. Science 171: 294-297.

Urata, Y., Parmelee, S.J., Agard, D.A., and Sedat, J.W. (1995): A three-dimensional structural dissection of Drosophila polytene chromosomes. J. Cell Biol. 131: 279—295.

Varmuza, S., Prideaux, V., Kothary, R., and Rossant, J. (1988): Polytene chromosomes in mouse trophoblast giant cells. Development 102: 127—134

Vig B.K. and Broccoli D. (1988): Sequence of Centromere separation: differential replication of pericentric heterochromatin in multicentric chromosomes. Chromosoma. 96(4): 311-317.

Wehr J.D. and Sheath R.G. (2003): Freshwater algae of North America: Ecology and Classification. Academic Press.

Whitefield W., Gonzalez G., MacDonald-Codina., Golver D., (1990): The A- and B-type cyclins are accumulated and destroyed in temporally distinct events that define separable phases of the G2-M transition. EMBO 9: 2563-572.

Zhang P and Spradling A.C. (1994): Insertional mutagenesis of Drosophila heterochromatin with single P-elements. P.N.A.S. 91: 3539-3543.

Zuchowski C.L. and Harford A.G. (1976): Unintegrated ribosomal genes in diplid and polytene tissues of Drosophila melanogaster. Chromosoma. 58(3): 235-246.

Zybina E.V. and Zybina T.G. (1996): Polytene chromosomes in mammalian cells. Int. Rev. Cytol. 165: 53-11.

This Page Is Intentionally Left Blank

Chapter 2: Incomplete replication generates somatic DNA alterations within Drosophila polytene salivary gland cells

Author Contribution

In this chapter is an edited reproduction of the original manuscript submitted of my published paper. This manuscript was prepared by the authors of this publication and is similar, but not identical to the version of this paper published and cited on the following page. Edits have been made to accommodate formatting requirements and additional comments by the author of this dissertation.

Citation

Yarosh W. and Spradling A.C. (2014): Incomplete replication generates somatic DNA alterations within Drosophila polytene salivary gland cells. Genes and Dev. 28: 1840-1855.

ACKNOWLEDGEMENTS

The authors are grateful to Allison Pinder for expert assistance with library construction and DNA sequencing. We thank Nick Ingolia and Fred Tan for valuable advice on sequence analysis. Steve DeLuca provided valuable assistance with chromatin analysis. A.C.S is an Investigator of the Howard Hughes Medical Institute. The authors declare no conflicts of interest.

Introduction

The idea that metazoan tissue cells contain identical genomes has long served as a convenient fiction appropriately termed "the dogma of DNA constancy" (Gilbert, 2011). In reality, despite highly faithful polymerases and repair systems, all organisms begin to sporadically accumulate DNA sequence alterations at low levels beginning with the first embryonic divisions (Reizel et al. 2013; Grandi and An, 2013; Kazazian, 2011). If replication is stressed (Lambert and Carr, 2012), or the cell cycle altered (Fox et al. 2010; Unhavaithaya and Orr-Weaver, 2012), greater levels of DNA changes may occur. Although well documented, these genome alterations have no known functional importance and are thought to be neutral or deleterious. At the other end of the spectrum, in relatively few organisms and cells, somatically programmed genomic changes generate useful differences. Eggshell genes are specifically amplified (Calvi and Spradling, 1999), antibody genes are productively rearranged (Alt et al. 2013), and whole ciliate genomes are re-engineered (Chalker and Yao, 2011).

The polytene cells of Dipterans such as Drosophila represent an intermediate case. During the growth of such cells via as many as ten consecutive endocycles (cell

cycles without cytokinesis), most euchromatic chromosome regions are fully replicated but pericentromeric genomic regions rich in satellite DNA sequences are not (Gall et al. 1971). In the best-studied system, the Drosophila larval salivary gland (Figure 1A), late-replicating euchromatic regions ("intercalary heterochromatin") also under-replicate to varying degrees (reviewed in Spradling and Orr-Weaver, 1987; Belyaeva et al. 2008). 30-52 under-replicated ("UR regions") 90-570 kb in length have been precisely mapped using DNA arrays (Belyakin et al. 2005; Nordman et al. 2011; Sher et al. 2012). These UR zones correspond closely to regions of repressive chromatin, sparse replication origins and mostly silent genes (Belyakin et al. 2005; Pindyurin et al. 2007; Filion et al. 2010; Nordman et al. 2011; Sher et al. 2012; Maksimov et al. 2013). The repressive chromatin state and late replication timing of UR regions is thought to be responsible for their susceptibility to incomplete replication.

The biological significance of under-replication has remained unclear. Most UR regions are the same in polytene fat body, midgut and salivary gland tissues, but a few show tissue-specificity suggestive of a developmental function; moreover, genes in UR regions in fat body are also more

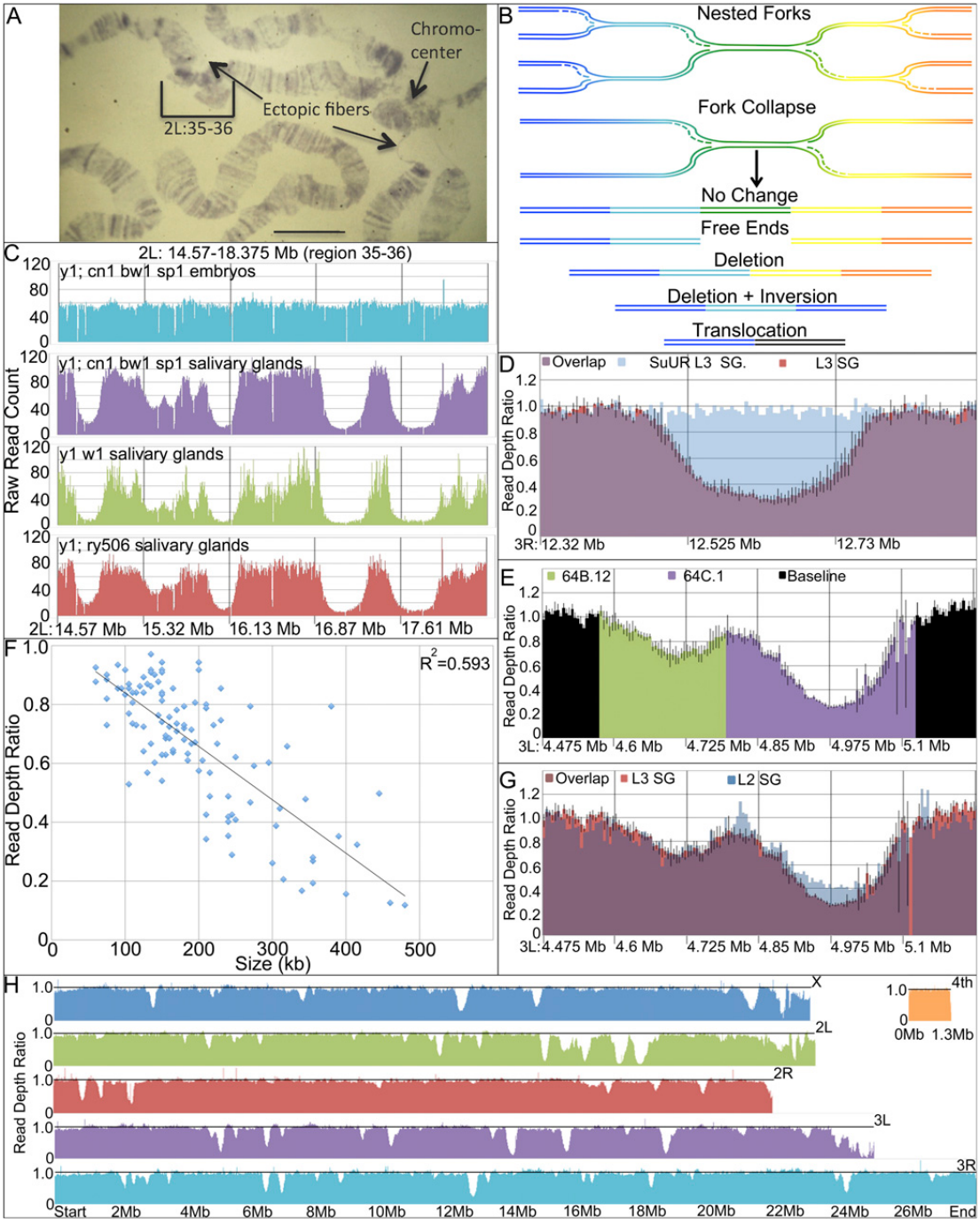


Figure 1. Detailed mapping of under-replication in larval salivary gland DNA by sequencing

Figure 1 continued from previous page

A) Drosophila larval polytene salivary gland chromosomes showing banded euchromatic arms (bracket: 2L:16.5-18.0 Mb, with ectopic fibers) and heterochromatic chromocenter (arrowhead). B) Models of under-replication. Stalled forks may be stable, forming inverse "onionskins." Alternatively, stalled forks may break and undergo repair, leading to novel DNA junctions and genomic alterations. C) Read counts in 2L:16.5-18.0 Mb, the same region shown in brackets in A. The read depth is uniform in embryo DNA, whereas multiple under-replicated (UR) regions are consistently found in the same locations in salivary gland DNA from three different strains. D) Average salivary gland/embryo read ratio from three separate experiments (orange) showing under-replication in UR89D.4. Bars = standard deviation. In *SuUR* mutants UR89D.4 is fully replicated (blue). E). Example of quantitation of UR regions 64B.12 and 64C.1. Bars = standard deviation. F) Under-replication in 2nd instar (blue) compared to 3rd instar salivary glands (orange). UR64B.12 appears similar, while UR64C.1 is less under-replicated in l2. G) Under-replication throughout the Drosophila genome based on read ratios (see D).

Text continued from page 44

frequently expressed (Nordman et al. 2011). Genetic evidence suggests that any such function is non-essential, however. A specific gene, *suppressor of underreplication (SuUR)*, encoding a novel protein, is required for differential replication in euchromatin but mutants are viable (Belyaeva et al. 1998). SuUR is found within many UR regions (Pindyurin et al. 2007; Nordman et al. 2011; Sher et al. 2012; Koryakov et al 2011) but is also distributed widely elsewhere in Drosophila chromatin (Filion et al. 2010; Maksimov et al. 2013). SuUR is proposed to slow the progress of replication forks preventing S phase completion in susceptible regions (Sher et al. 2012; Kolesnikova et al. 2013). Failure to complete replication might cause a mitotically proliferating cell to undergo apoptosis, however, many endocycling cells down-regulate the normal apoptotic response to unrepaired DNA damage (Mehotra et al. 2008).

A better understanding of the molecular consequences of under-replication would likely help reveal its significance. If stalled replication forks remain stable, under-replicated domains would contain inverse DNA "onionskins" linked by the forks (Laird, 1980; Sher et al. 2012; see Figure 1B). In contrast, if forks undergo

breakage, then free ends would be produced, which if repaired would cause DNA rearrangements (Spradling, 1993; Leach et al. 2000; Andreyeva et al. 2008). Previous searches failed to detect stable replication forks in an under-replicated region (Glaser et al. 1992). However, novel DNA bands were observed in Southern blots of DNA from polytene tissues, consistent with DNA breakage (Glaser et al. 1992; Spradling, 1993; Glaser et al. 1997; Leach et al. 2000). Distinctive features of polytene chromosome structure such the chromocenter, and ectopic fibers such as those consistently observed in polytene region 35-36, might also be explained by a high level of breakage and repair (Figure 1A). The genetic phenomenon of position-effect variegation has also been ascribed to DNA alterations (Karpen and Spradling, 1990). However, most researchers have rejected the idea of somatic DNA instability (Ahmad and Golic, 1996).

Here we analyze polytene DNA using high throughput sequencing and show that DNA alterations are generated at many sites throughout the genomes of salivary gland and ovarian cells. DNA deletions 10-500 kb in size are found throughout 112 UR zones comprising 19% of salivary gland euchromatin, but are rare within fully replicated regions or in the corresponding regions of early embryonic diploid

cells. Thus, during polytenization unfinished replication forks break and efficiently rejoin to nearby free ends. An even higher level of under-replication takes place in heterochromatic regions whose repetitive sequences prevented detailed analysis using our methods. Our results show that somatic DNA instability is a widespread feature of polyploid Drosophila cells. The significance of somatic DNA modification for chromosome structure, position effect variegation and developmental function deserves further study.

RESULTS

Deep sequence analysis of under-replication

DNA from early Drosophila embryos, whose cells are predominantly diploid, and from third instar larval salivary glands, which are comprised mostly of highly polyploid cells was prepared in order to study under-replication by deep sequencing (Table 1). The *y; cn bw sp* reference strain used to determine the Drosophila genome sequence was employed in order to minimize alignment ambiguities (Adams et al. 2000) as well as the *y w* and *y ; ry[506]* strains. Each DNA preparation was sheared, and libraries were prepared and subjected to paired-end

Genotype	Tissue	Name	Reads	Alignment
OreR P2	S10-14 follicle	S1014F	86327105	CASAVA 1.7
y w	L3 salivary gland	ywSG	87826055	CASAVA 1.7
y ; ry[506]	L3 salivary gland	yrySG	134328874	CASAVA 1.8
y ; cn bw sp	L3 salivary gland	RefSG1	125797138	BWA
y ; cn bw sp	L3 salivary gland	RefSG2	176722412	BWA
y ; cn bw sp	L3 salivary gland	RefSG3	137202148	BWA
y ; cn bw sp	0.5-2.5 hr embryo	RefEmb1	163631228	BWA
y ; cn bw sp	0.5-2.5 hr embryo	RefEmb2	138242266	BWA
y ; cn bw sp	0.5-2.5 hr embryo	RefEmb3	101037972	BWA
y ; cn bw sp	Whole ovary	RefOV	142944736	BWA
SuUR	L3 salivary gland	SuURSG	72043065	BWA
SuUR	0.5-2.5 hr embryo	SuUREm	65701687	BWA
y ; cn bw sp	L2 salivary gland	L2SG	73183306	CASAVA 1.8

Table 1: DNAs analyzed. The DNAs that were sequenced and analyzed as part of this study. The genotype and tissue source is given along with a short abbreviation for each preparation as used in the paper. The number of raw reads and alignment software used is also shown.

Text continued from page 49

sequencing with 100 bp reads on an Ilumina HiSeq2000. Sequences were aligned to the Drosophila genome R5.33 using ELAND and BWA software (see Methods).

We tested the utility of sequencing for analyzing changes in genomic copy number by examining the behavior of heterochromatic sequences. The severe under-replication of heterochromatin in salivary gland DNA was evident from the fraction of reads that aligned to heterochromatic vs. euchromatic zones of the genome. In embryo DNA 35.6% of read pairs mapped to heterochromatic regions, whereas only 2.2% of pairs from the salivary gland aligned to heterochromatin (see Table 2). Raw sequencing reads were queried to estimate sequence under-replication "digitally" based on read frequencies, and compared with previous "analog" assessments based on nuclei acid hybridization. 0.15%-1.2% of reads from embryo DNA but only 0.005%-0.05% of the reads from salivary gland DNA were homologous to individual satellite DNAs of *Drosophila melanogaster*. Thus, satellite DNAs appear to be under-replicated about 30 fold (~ 5 rounds of replication) during salivary gland development, a somewhat smaller degree of under-replication than previously reported in *D. melanogaster* (Rudkin, 1969)

Contig	chromatin	size	2-end	1-end pr.
Uextra	Het	29004656	2235828	1089732
3R	Eu	27905053	21270740	345851
3L	Eu	24543557	18649653	392509
2L	Eu	23011544	17713153	341295
X	Eu	22422827	12017915	274407
2R	Eu	21146708	15915102	331272
U	Het	10049037	23164607	1079205
2RHet	Het	3288761	1145922	164650
3LHet	Het	2555491	1062195	151592
3RHet	Het	2517507	1016923	154613
4	Eu	1351857	1129429	38039
2LHet	Het	368872	119684	20764
YHet	Het	347038	33492	5192
XHet	Het	204112	72914	5354
Unaligned		0	0	43689180
Subtotal			115547557	4394475
Total			163631212	

Alignment statistics for RefEmb1 (see legend below for details).

Contig	chromatin	size	2-end	1-end
Uextra	Het	29004656	136994	78190
3R	Eu	27905053	27380069	297306
3L	Eu	24543557	22844698	303093
2L	Eu	23011544	20960512	280590
X	Eu	22422827	17471119	274157
2R	Eu	21146708	20606327	304311
U	Het	10049037	619200	55832
2RHet	Het	3288761	543610	72680
3LHet	Het	2555491	452236	53455

Table 2 continued from previous page.

3RHet	Het	2517507	266943	32868
4	Eu	1351857	1306918	34063
2LHet	Het	368872	73569	10033
YHet	Het	347038	757	201
XHet	Het	204112	66025	2864
Unaligned		0	0	11268502
Subtotal			112728977	1799643
Total			125797122	

Table 2. Emb1 and SG1 alignment summary Alignment statistics for RefSG1. Drosophila genomic contigs from genome R5.33 are shown (Contig) and whether they correspond to euchromatin (Eu) or heterochromatin (Het) is indicated (chromatin). The number of read pairs in which both reads (2-end) or 1 read (1-end) aligned to the indicated genomic contigs is indicated. The % of read pairs aligning to heterochromatin is calculated by summing all read pairs that aligned to heterochromatic contigs and dividing by total aligned read pairs. (Both 2-end and 1-end pairs are considered to be aligned). A much smaller fraction of reads aligns to heterochromatin in DNA from salivary gland (RefSG1) than from early embryos (RefEmb1).

Text continued from page 51

or *D. virilis* Gall et al. 1971). The read frequency of rDNA sequences in salivary gland averaged 24 +/- 5.9% that of embryo (N=3) (Table 3), consistent with the four-fold under-representation previously reported (Spear and Gall, 1971). 5S rDNA, which is encoded in a separate locus, replicated fully (Table 3) as expected (Hammond and Laird, 1985).

More than 100 chromosome regions under-replicate in larval salivary glands

Plotting the frequency of reads along the euchromatic Drosophila genome sequence potentially provides a highly sensitive measure of replication uniformity. When DNA was sheared but not narrowly sized prior to library construction, read frequency from embryo DNA was highly uniform along the five major chromosome arms (Figure 1C; see Methods). In contrast, plotting the read frequency in salivary gland DNA revealed many chromosome regions where read frequency declines smoothly over a distance of 50-100 kb and then increases back to the genome average in a strain independent manner (Figure 1C). The most strongly affected zones correspond to the major UR regions mapped previously (Belyakin et al. 2005; Sher et al. 2012;), such

Table 3 Sequence under-replication of repetitive DNAs

DNA	Reads x106	AAGAG	AATAT	AAGAC	AATAC	AAGAGAG	rDNA	5S DNA
Emb1	**163.0**	**1020728**	**858187**	**163157**	**127766**	**143706**	**8651**	**1068**
		1.252	1.053	0.200	0.157	0.176	0.011	0.0013
Emb2	**138.2**	**659821**	**20861**	**176999**	**115597**	**152489**	**8649**	**1055**
		0.955	0.030	0.256	0.167	0.221	0.013	0.0015
Emb3	**101.0**	**443433**	**15107**	**111663**	**68824**	**102870**	**6639**	**712**
		0.878	0.030	0.221	0.136	0.204	0.013	0.0014
SG1	**125.8**	**31526**	**14341**	**12570**	**3356**	**5405**	**1497**	**1652**
		0.050	0.023	0.020	0.005	0.009	0.002	0.0026
SG2	**176.7**	**13337**	**3839**	**18810**	**5184**	**2816**	**2378**	**1908**
		0.015	0.004	0.021	0.006	0.003	0.003	0.0022
SG3	**137.2**	**9948**	**2682**	**14286**	**4282**	**2861**	**1831**	**1474**
		0.015	0.004	0.021	0.006	0.004	0.003	0.0021
O	**142.9**	**280422**	**78284**	**42027**	**525**	**40076**	**5757**	**1293**
		0.392	0.110	0.059	0.001	0.056	0.008	0.0018
SG/ Emb								
1		0.040	0.022	0.100	0.034	0.049	0.224	2.004
2		0.016	0.144	0.083	0.035	0.014	0.215	1.414
3		0.017	0.131	0.094	0.046	0.020	0.203	1.524

Table 3 Sequence under-replication of repetitive DNAs

The number of raw sequence reads generated from the indicated DNAs that match 15bp probe sequences from the following sources: satellite DNAs, 18S rDNA or 5S DNA. The percentage of total reads is shown below, and the last three lines display abundance ratios between salivary gland and embryo DNAs.

Text continued from page 54

as the UR zones in chromosome 2L regions 35 and 36 (Figure 1C).

The UR zones were characterized more accurately by averaging reads in 5kb windows across the genome and normalizing salivary gland reads to embryo reads in each window to minimize perturbations caused by the presence of repetitive DNA. Normalized 5kb read values were calculated based on three separate experiments, each involving one preparation of embryo and salivary gland DNA that was processed separately to control for effects particular to any given DNA or library preparation. The average ratios and standard deviations were plotted to determine the replication profiles (Figure 1D, E). This approach revealed that more euchromatic UR regions exist than described previously, but that most of them show only low levels of under-replication. Overall, we defined 112 euchromatic regions 60-480 kb in size that were consistently under-replicated in salivary gland DNA (Table 4). Twenty-five regions were between twofold and 8.46-fold under-replicated, while the remaining 87 zones were reduced 1.06-1.99 fold. The level of under-replication correlated with the size of the region (Figure 1F).

Together, the UR domains account for 21.76 Mb / 115.7 Mb = 19% of the euchromatic genome.

Previous studies have shown that euchromatic under-replication is greatly reduced or absent in *SuUR* mutant salivary glands (Belyaeva et al. 1998; Sher et al 2012). When salivary gland and embryo DNA from $SuUR^{-/-}$ animals was examined, nearly all the UR regions, including those with low UR values, were greatly attenuated, further supporting their validity (Figure 1D; Table 4). We also found (Figure S1), as previously reported for strong UR regions (Sher et al. 2012), that all the UR regions corresponded closely to domains of repressive chromatin as defined in genomic studies (Karchenko et al. 2012). To investigate when under-replication occurs during development we analyzed second instar larval (L2) salivary glands, which have completed about 7 endocycles. Under- replication in most regions was already complete by the second instar (Figure 1F; Table 4). However, the most strongly under-replicated regions were exceptional; in the second instar these regions showed less under-replication than in mature third instar (L3) glands. Rather than supporting developmental regulation, these are the results expected if replication failure has a constant probability characteristic of each UR region during each endocycle.

UR	Mb	kb	C	L3SG	L2SG	SUUR	Ovary	DfSG	DfEm	Genes
X										
3C.3	2.720–2.810	90	b	1.17	1.20	1.03	1.06	15	0	
3C.6	2.810–3.060	250	b	2.45	1.21	1.00	1.07	39	2	kirre, Rst
4C.3	4.280–4.340	60	g	1.14	1.02	1.10	1.05	5	0	bi
4D.3	4.620–4.780	160	b	1.38	1.24	1.09	1.19	21	0	Proc-R
4E.1	4.850–4.960	110	b	1.24	1.25	1.11	1.33	8	0	ovo,
7B.3	7.210–7.485	275	b g	1.69	1.71	1.03	1.19	26	0	CG1677, ct
7C.1	7.650–7.725	75	b	1.37	1.13	1.00	1.08	9	1	Ir7b, Ir7a
8E.1	9.290–9.395	105	r	1.17	1.16	1.01	1.05	12	0	Megalin
8F.10	9.570–9.675	105	g	1.15	1.09	1.06	1.15	14	1	btd
9A.2	9.675–9.880	205	b	1.64	0.90	0.98	0.97	17	0	CG32698
10B.2	11.100–11.175	75	b	1.22	1.22	1.02	1.13	9	0	CG42683
11A.6	11.900–12.285	415	u r	3.09	2.19	1.06	1.31	48	0	Ten-a
12A.2	13.300–13.400	100	b	1.09	1.06	1.04	1.08	9	0	CG32635
12E.2	13.900–14.050	150	b	1.19	1.07	1.05	1.16	6	0	Ste
12E.8	14.15–14.465	310	b	2.24	1.83	1.04	1.29	36	4	dpr8
13B.1	15.045–15.120	75	b	1.11	1.09	1.00	1.09	9	0	CG9095
14B.3	16.000–16.150	150	g	1.23	1.17	1.06	1.14	14	0	disco, disco-r
16F.3	17.825–17950	125	u	1.12	1.09	0.94	1.06	10	0	Sh
17A.1	18.000–18.200	200	u g	1.06	1.06	1.03	1.06	10	1	Frq, upd
19A.4	19.795–20.005	210	u	1.25	1.09	0.98	1.13	16	1	Dop2R
19D.3	20.350–20.485	135	b	1.50	1.26	1.01	1.06	22	1	RunxA, RunxB
19E.5	20.485–20.875	390	b	2.84	2.31	1.07	1.32	49	2	shakB, Npc1b
20A.1	21.335–21.460	105	u	1.89	1.95	0.97	1.21	8	0	CG42343
2L										
21E.2	0.630–0.745	115	b	1.19	1.16	1.07	1.06	10	0	ds
22A.1	1.2701–1.500	230	b	1.17	1.18	1.08	1.39	16	3	lea,robo 3
23A.2	2.605–2.740	135	b	1.09	1.10	1.00	1.02	10	0	CG31690

Table 4 continued from previous page

24D.1	3.905—4.015	110	b	1.36	1.42	1.02	1.09	34	1	fred, ed
25A.3	4.565—4.780	215	b	1.76	1.58	1.02	1.24	30	1	CG15630, dp
25E.1	5.370—5.530	160	g	1.32	1.45	1.08	1.12	8	0	H15,mid
25F.1	5.555—5.690	135	b	1.16	1.21	1.05	1.02	12	0	CG14010, CG31646
26C.1	6.150—6.305	155	b	1.46	1.39	1.08	1.14	25	2	Ddr, CG34380
29F.7	9.005—9.115	110	b	1.19	1.31	1.02	1.05	17	1	CG31708
32A.1	10.530—10.685	155	b	1.59	1.20	1.06	1.12	24	0	Trim9,
32F.1	11.305—11.475	170	g	1.28	1.46	1.04	1.17	18	2	salr, salm
32F.3	11.525—11.765	240	b	2.83	2.10	1.02	1.22	57	0	kek2
33E.1	12.200—12.315	115	b	1.37	1.33	1.00	1.04	33	1	aret
34A.1	12.750—12.950	200	b	1.74	1.66	1.02	1.07	27	3	kek1, ACXC
34F.2	13.900—14.100	200	b	1.09	1.05	1.01	1.13	12	0	nAcRa
35B.6	14.670—14.985	315	b	4.86	2.58	1.03	1.34	65	2	CG42313
35D.1	15.240—15.510	270	b	2.14	2.10	1.08	1.24	46	0	
35D.3	15.510—15.660	150	b	1.85	1.73	1.05	1.07	49	7	kek3
35D.4	15.770—15.900	130	b	1.76	1.64	1.08	1.21	55	5	CG13243
35E.2	15.905—16.245	340	b	5.98	4.26	1.07	1.24	78	4	beat-Ia,b,c
35F.12	16.360—16.450	90	bg	1.17	1.16	1.05	1.17	7	0	jhamt, CG5888
36B.3	16.900—17.360	460	b	7.94	5.10	1.07	1.57	56	1	beat-IIIa,b,c
36E.1	17.500—17.980	480	b	8.46	4.01	1.31	1.58	103	3	CadN, CadN2
36E.3	17.980—18.170	190	b	1.40	1.39	1.05	1.14	35	2	rdo
36F.1	18.170—18.275	105	b	1.30	1.52	1.05	1.19	24	1	CG42750
37D.1	19.210—19.345	135	b	1.10	1.13	1.08	1.24	9	0	
38C.3	20.055—20.285	230	u	1.34	1.30	1.07	1.17	10	0	
38D.1	20.485—20.630	145	b	1.14	1.18	0.97	1.04	14	0	piRNA5
40D.1	21.770—22.115	345	u	2.09	2.39	1.13	1.24	52	0	tsh, cg31612
2R										
41F.1	1.270—1.485	215	u	2.05	1.58	1.02	1.09	9	2	CG42345,

Table 4 continued from previous page

42A.1	1.700—1.825	125	u	1.56	1.38	1.04	1.08	8	0	dpr12
42A.14	2.140—2.385	245	u	3.46	3.04	1.03	1.17	1	0	piRNA1
50A.1	9.150—9.300	150	b	1.09	1.07	1.03	1.06	9	1	Dh31-R1
50C.1	9.480—9.650	170	b	1.36	1.12	1.08	1.08	20	1	fas
53C.5	12.250—12.380	130	b	1.15	1.16	1.08	1.03	18	1	sema2a, sema2b
55A.3	13.805—13.950	150	b	1.09	1.15	0.96	1.02	15	1	dpr13, CG34386
56F.1	15.700—16.080	380	b	1.26	1.23	1.11	1.23	25	0	18w,
57A.1	16.175—16.445	270	b	1.26	1.23	1.06	1.36	33	1	CG12484, Toll-7
58A.1	17.585—17.835	250	b	1.61	1.46	1.05	1.10	27	1	Fili
59D.3	18.930—19.225	295	b	1.66	1.56	1.06	1.17	28	3	CG34371
60F.1	20.900—21.140	240	gb	2.05	1.93	1.05	1.59	28	1	CG9380, lov
3L										
63A.1	2.855—2.915	60	b	1.08	1.09	0.98	1.10	6	0	Shab
64B.12	4.605—4.795	190	b	1.50	1.38	1.06	1.20	46	0	axo, Gef64C
64C.1	4.795—5.095	300	b	3.83	2.36	1.05	1.23	85	1	Con
64D.1	5.350—5.500	150	b	1.59	1.66	0.98	1.07	27	2	CG34391
65A.14	6.260—6.470	210	b	2.92	2.01	1.03	1.17	38	0	Or65, CG42747
65C.5	6.75—6.905	155	g	1.44	1.38	1.07	1.09	19	2	vvl
67A.4	9.125—9.310	185	b	1.64	1.26	1.04	1.02	25	2	Glu-RIB
67D.2	9.960—10.205	245	e	2.35	1.77	1.01	1.09	52	1	dpr6, dpr10
67D.11	10.20—10.350	145	rb	1.31	1.44	1.04	1.18	31	0	Or67
67F.1	10.920—10.995	75	b	1.13	1.13	1.05	1.09	6	2	klu
70A.2	13.010—13.235	225	b	1.50	1.29	1.06	1.11	46	0	caps
70C.2	13.480—13.880	400	b	6.41	3.28	1.03	1.41	64	1	bru-3, dysc
71A.1	14.835—15.045	210	b	1.27	1.58	1.09	1.32	4	1	CG17839
71C.2	15.125—15.480	355	b	3.73	2.42	1.07	1.16	58	0	Tollo, Toll-6, Best4
74A.1	17.075—17.225	150	b	1.35	1.27	1.01	1.04	32	1	Rbp6

Table 4 continued from previous page

75C.1	18.105—18.460	355	rb	5.18	3.07	1.06	1.30	89	0	grim, rpr, CheA75A
75D.1	18.460—18.630	170	b	1.41	1.20	1.03	1.10	15	0	AstC-R2
77E.1	20.535—20.715	180	g	1.39	1.63	1.07	1.13	31	1	knrl, kni
79E.3	22.345—22.665	320	b	1.52	1.47	1.04	1.21	27	1	Ten-m
3R										
83E.1	1.725—2.170	445	b	2.01	1.75	1.04	1.30	34	1	CG34113, Osi, cluster
83F.1	2.305—2.470	165	b	1.57	1.34	1.01	1.12	39	0	dpr11
84A.5	2.525—2.690	165	g	1.17	1.20	1.07	1.23	9	1	pb, zen2, Dfd
84B.1	2.690—2.875	185	g	1.58	1.45	1.09	1.25	23	1	Scr, Antp
84D.10	3.380—3.545	165	b	1.55	1.40	0.99	1.18	38	1	Cg34127, Nlg1
85A.1	4.200—4.350	150	b	1.10	1.07	1.06	1.06	13	2	Or85b, cluster
86C.1	6.265—6.505	240	g	2.39	1.57	1.15	1.32	51	0	hth
86D.1	6.735—6.975	240	b	2.49	1.68	1.02	1.09	62	1	CG34114
87A.9	7.865—7.990	125	rb	1.38	1.30	1.04	1.10	24	0	dpr15, dpr 17
87D.1	8.555—8.750	195	b	1.49	1.60	0.99	1.08	35	1	beat-Va,b,c
87F.1	9.260—9.400	140	b	1.11	1.06	1.03	1.03	13	0	CG14372
88D.9	10.750—10.845	95	e	1.17	1.25	0.99	1.02	20	1	dpr9, CG14861
89A.1	11.395—11.590	195	gb	1.24	1.20	1.05	1.14	11	1	pxb, Fe/S, cluster
89E.1	12.465—12.820	355	g	3.57	2.28	1.11	1.26	76	2	BX-C
90A.1	12.995—13.175	180	b	1.47	1.44	1.00	1.05	40	0	beat-Iia,b
92D.1	15.905—16.045	140	b	1.66	1.43	0.97	1.10	37	1	Nlg4, CG5060
92E.8	1.6160—16.320	160	b	1.46	1.21	1.04	1.03	29	1	Gfrl, Ir92a
94A.2	17.920—18.130	210	b	2.35	1.67	1.08	1.09	36	1	SKIP, Ir, cluster
94D.3	18.655—18.755	100	b	1.20	1.09	1.02	1.04	13	0	klg
95A.1	19.210—19.390	180	b	1.26	1.18	1.00	1.03	21	1	beat-IV, Ir-94, cluster

Table 4 continued from previous page

96A.1	20.195–20.330	135	b	1.03	1.03	1.03	1.02	13	0	nAcRa-96Aa,b
97F.6	23.150–23.300	150	b	1.06	0.99	1.05	1.05	11	1	side
98B.3	23.545–23.725	180	b	1.37	1.22	1.02	1.04	30	0	CG34353
98C.2	23.790–24.095	305	b	2.58	2.09	1.03	1.20	74	4	CG34362, CG34354
98D.1	24.165–24.255	90	b	1.07	1.06	1.01	1.04	23	1	beat-VI
99A.7	25.150–25300	150	b	1.20	0.95	0.98	0.97	11	2	Ptp99A
100A.2	26.405–26.625	220	br	1.37	1.18	1.07	1.06	14	0	zfh1, Pka-C2
100B.1	26.715–26.840	125	gb	1.19	1.16	1.06	1.08	17	2	Ptx1, 5-HT7
100B.4	26.860–26.980	120	gb	1.16	1.05	1.02	1.06	11	1	sox100B, Gycb100B

Table 4: Regions of Under Replication. UR regions are listed showing their genomic coordinates (Mb), size (kb), chromatin type (C): state 1-5 red (r); state 6: green (g); state 7-8: blue (u), state 9: black (b). fold underreplication (UR) in L3 or L2 salivary gland (SG), SuUR L3 SG, Ovary, deletions (Df) recovered in L3 salivary gland (SG) or early embryo (Emb), and selected genes within the UR.

Text continued from page 57

Replication appeared to be uniform across the genome, except within the 112 UR regions (Figure 1G). The measured copy number values indicate that even within UR regions complete replication is the norm. In only a few regions, such as UR 36B.3, UR36E.1 and UR70C.2, are more than half of the wild type strands mutated over the course of 10 doublings. In most UR regions, infrequent replication failure would be sufficient to explain their UR values (less than 2). Thus, copy number changes during salivary gland development arise stochastically in multiple regions by relatively low absolute levels of incomplete replication.

Ovary DNA also contains under-replicated regions

Many larval and adult Drosophila tissues in addition to salivary gland are made up predominantly of polyploid cells that contain polytene chromosomes and under-replicate heterochromatin (Ashburner, 1972; Spradling and Orr-Weaver, 1987). Genomic analyses have shown that at least two such tissues, larval fat body and larval midgut, under-replicate most of the same euchromatic regions as salivary gland (Nordman et al. 2011). We sequenced DNA prepared from ovaries, which derives predominantly from polytene nurse and follicle cells, to examine the euchromatic regions that

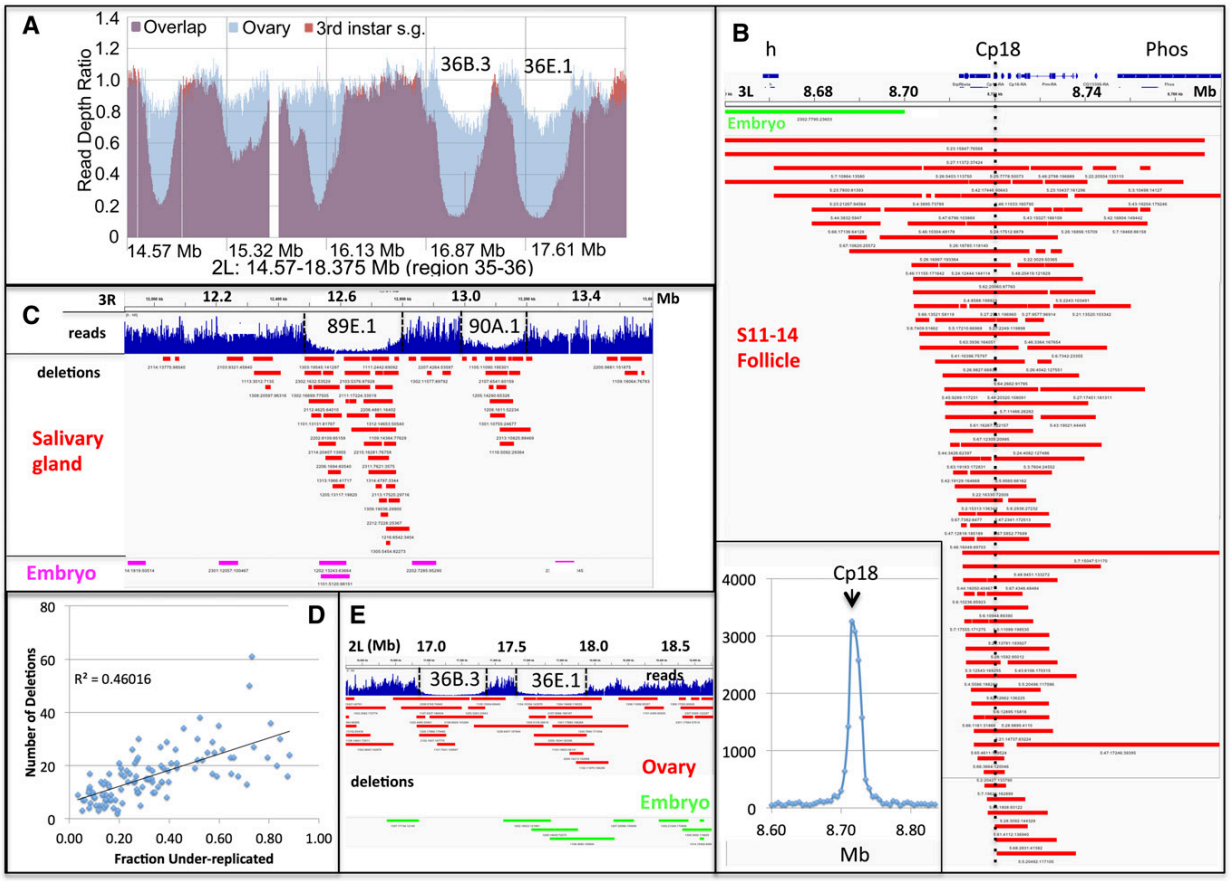


Figure 2. Replication fork instability leads to DNA deletions

A) Normalized read frequency from ovary DNA (green) in region 2L:14.5-18.3 Mb contains many of the same URs as salivary gland DNA (brown). B) Map of deletions surrounding the amplified 66D chorion gene cluster at 3L:8.66-8.76 Mb. Deletions are abundant in DNA from amplification stage follicles (red) and are centered on the major origins located near Cp18 (dashed line), but are rare in embryo DNA (green). C) Chromatin types from Bc3 cultured cells are labeled according to Karchenko et al.

Figure 2 continued from previous page.

(2013), and colored in groups for convenience: states 1-5: red (r); state 6: green (g); state 7-8: blue (e), state 9: black. Salivary gland DNA contains many more deletions 10-500kb in length (red) than embryo DNA (green). D) Deletions are enriched in salivary gland DNA and the excess deletions in UR regions are proportional to the degree of under-replication. E) Ovary DNA (red) also contains more deletions in UR regions than in embryo DNA (green).

Text continued from page 63

under-replicate in adult polytene cells. Most UR regions that under-replicate strongly in larval salivary glands duplicated incompletely during ovary development, although the level of reduction in ovary DNA was much less, in part due to the presence of diploid ovarian cells (Figure 2A, Table 4). Thus, the UR regions defined for salivary gland are likely to be similar in a wide range of polyploid cells, including nurse and follicle cells.

Replication forks are unstable during chorion gene amplification in ovarian follicle cells

In order to distinguish whether polytene DNA contains stalled forks or has undergone breakage/repair, methods are needed that can identify rare molecules with novel junctions. To assess the ability of paired end sequencing to detect rare products of replication fork instability we initially investigated dense zones of replication forks that are generated during chorion gene amplification. Amplification generates a high density of replication forks because multiple rounds of replication initiate at just a few genomic locations during a final S-like phase in stage 10B of oogenesis (Calvi and Spradling, 1996). Subsequently, during stages 11-14, these forks continue to elongate moving away from the initiation region on each

side (Claycomb et al. 2002; Nordman and Orr-Weaver, 2012). If replication forks break and are repaired in vivo, novel junctions will be generated that could be detected by paired end sequencing.

We sequenced DNA from stage 11-14 follicles and looked for anomalous read pairs, i.e. those whose component reads align at sites incompatible with normal sheared DNA. The read profile revealed the dramatic nature of amplification. For example, in the largest amplified domain on chromosome 3L, comprising about 100 kb centered on Cp18 and 3 other chorion genes, the read profile increased in exactly the location expected from previous studies, with a peak value 47 times higher than the average in unamplified regions (Figure 2B, inset). More than 70 anomalous read pairs were identified in the amplified region, all indicative of deletions centered around Cp18, whereas only 1 deletion was found in embryo DNA (Figure 2B). Similar deletions (but less enriched) were observed around Cp18 using the ovary DNA sample, and at the other major site of gene amplification on the X chromosome (Figure 3). Thus, some replication forks, about 2% (Methods), break and undergo repair to form deletions during amplification, and these rare events can be detected by paired-end sequencing. The major origins used during amplification are located near

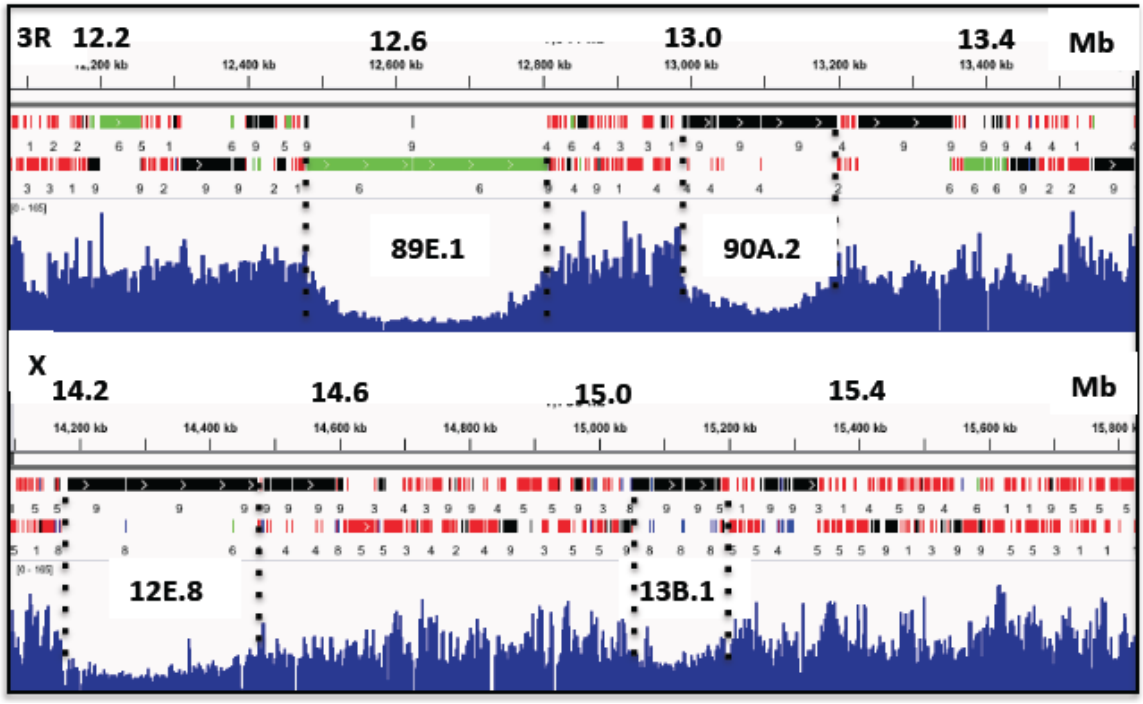


Figure 3. UR regions correspond closely with chromatin domains. The figure shows two genomic segments, 3R:12,100,000-13,600,000, and X:14,100,000-15,800,000, that each contain two UR regions (labeled). Within each displayed region, chromatin types (1-9) according to Karchenko et al. (2012) from Drosophila S2 cells are shown above a read track from L3 salivary gland. Chromatin subregions were colored as follows (1-5: red; 6: green; 7-8: blue; 9: black). Dashed lines indicate the boundaries of each domain. The correspondence between chromatin types and UR regions is clear over a wide range of under-replication levels: 89E.1 (3.57), 90A.1 (1.47), 12E.8 (2.24), 13B.2 (1.11); ( ) = UR value from Table 4.

Text continued from page 67

Cp18, so the presence of small deletions flanking this region suggests that some forks stall shortly after initiation, break and are repaired to other broken ends. These DNA alterations were not observed in earlier studies of chorion gene amplification, which illustrates the difficulty of detecting rare DNA derivatives that differ from each other.

Anomalous read pairs identify a class of deletions enriched in salivary gland DNA

With this encouragement, we took the same approach to look for DNA alterations generated during under-replication in salivary glands. Following alignment of paired-end sequences from salivary gland DNA to the Drosophila genome, we first analyzed anomalous pairs. Ideally, these pairs come from reads in which the un-sequenced center of the fragment contains a deletion breakpoint. However, a background of mis-leading anomalous read pairs will also be generated when reads are mis-aligned to the genome due to the presence of local repeats such as transposons or duplicated genes. Additionally, hybrid DNAs generated by the ligation of unrelated fragments during library preparation will also produce mis-leading anomalous pairs. Random ligation will generate “translocations” and large

"deletions" preferentially, since the chance two randomly joined fragments come from nearly the same chromosome region is less likely. If predicted DNA alterations truly result from DNA under-replication rather than methodological artifacts, then they should be enriched in salivary gland DNA compared to embryo DNA, and in UR regions compared to normal regions. We initially focused on deletions in the size range from 10kb to 500 kb, since 10kb is large enough to exclude transposon polymorphisms, and 500 kb is the upper limit of the measured UR size.

We identified all anomalous read pairs predicting 10-500 kb deletions among all salivary gland and embryo read pairs from the three replicate experiments using the reference strain. Pairs in which the sequence quality of one of the reads was questionable were excluded. We also removed read pairs with identical reads since they are a product of PCR amplification, and read pairs aligned within heterochromatin, including repetitive, unmapped portions of the genome (chrU and chrUextra). It was important to align against these sequences initially, however, to prevent matching reads from being force-aligned elsewhere in the genome.

The remaining read pairs were examined to determine if they might be related to under-replication. Salivary gland

DNA contained more than three times as many deletion read pairs as embryo DNA within euchromatin as a whole (Table 2). However, when the location of the salivary gland deletions was plotted, they showed only a slight specificity for UR regions (not shown). Because large deletions are more likely to be caused by random ligation, we tried plotting only those deletions with predicted sizes between 10kb and 100kb. Anomalous pairs in this size regime were fourfold enriched in salivary gland vs embryo DNA (Table 2) and the deletions they specify clustered within UR regions (Figure 2C). Significantly, the number of excess deletions in salivary gland DNA compared to embryo DNA in the UR regions correlated with the degree of under-replication (Figure 2D).

Anomalous pairs from ovary DNA also predicted a large excess of deletions compared to a corresponding number of reads from early embryos. However, the distribution of deletions differed from that seen with the salivary gland. Some strong UR regions such as 36B.3 and 36E.1 that under-replicate in the ovary (Table 4) contained more deletions in ovarian than in embryo DNA (Figure 2E). However, the overall level of enrichment in ovary URs (Table 4) compared to non-UR regions was modest. A strong peak was seen

DNA	Reads xE06	Eu. Depth xE06	Pre-dicted	Anom. Pairs 10-500kb	Anom. Pairs 10-100kb	Partially mapped reads	BP reads	In UR
SG1	125.8	96.08	3562	1682	824	133729	1375	1048
SG2	176.7	83.66	3102	6086	2075	121389	1328	1025
SG3	137.2	57.74	2141	5829	1638	93701	891	738
Emb1	163.6	74.4		410	252	220495	297	80
Emb2	119.9	34.01		2477	596	116271	55	15
Emb3	101	12.57		879	288	85336	36	8
total SG	439.7	358.5	8805	13597	4537	348819	3594	2811
total Emb	384.5	121		3766	1136	422102	388	112

Table 2. Identification of deletions by paired-end sequencing. Numbers of read pairs of various types from the indicated DNA preparations. Euchromatic (Eu.) read depth is after correcting for pairs that failed to align, or aligned to heterochromatin, and for PCR duplicates. Predicted: The approximate number of deletions required to completely account for Table 4 URs at this read depth. Observed anomalous (Anom.) pairs specifying deletions 10kb-500kb in size, or 10-100kb in size. The starting number of pairs with a partially mapped read, is compared to identified breakpoint (BP)-containing reads, and BP reads overlapping a Table 4 UR region.

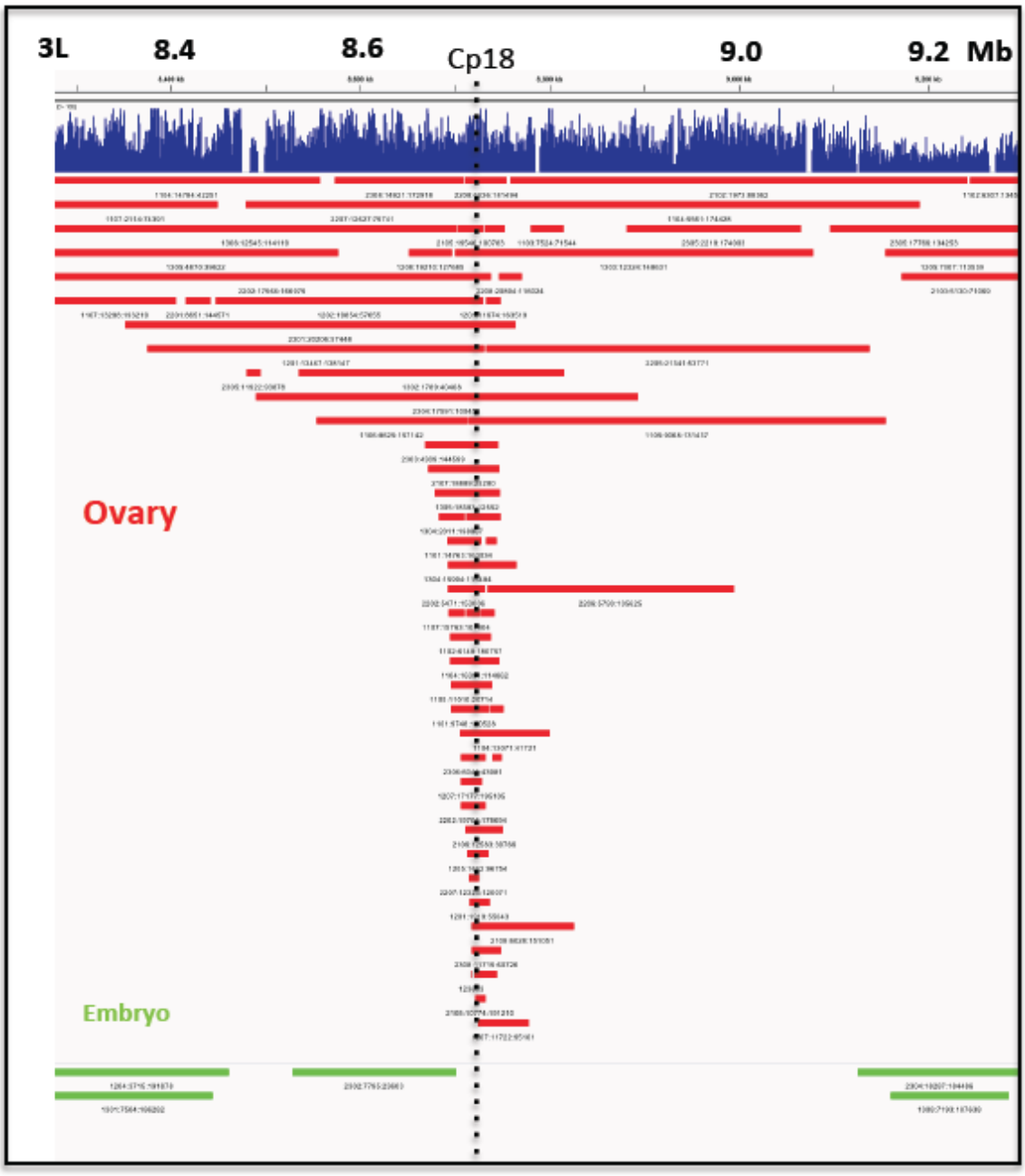


Figure 4. A large excess of deletions defined by ovary DNA anomalous pairs is found in the vicinity of amplified chorion genes. Genomic region 3L: 8,200,000-9,300,000 is displayed along with a read track from L3 salivary gland DNA to indicate the location of UR regions. Below the read track, deletions in this region defined by anomalous pairs

Figure 4 continued from previous page.

in Ref ovary DNA (red) and Emb1 (green) are plotted. Although, no UR is present, an excess of deletions is evident, like in stage10-14 follicle DNA (Figure 2B). The deletions are clustered around the Cp18 gene (dashed line), which marks the location of the major amplification origins.

Text continued from page 71

around Cp18 (Figure 4), documenting that biologically significant deletions were being observed.

These results provide strong evidence that under-replication generates somatic deletions in UR regions due to fork breakage and repair. However, the deletions defined by anomalous read pairs did not appear to reveal the entire distribution of rearrangements associated with under-replication. Reads larger in the 100kb -500kb range showed less specificity, suggesting that larger deletions result from random ligation, or an unknown biological process. The number of such deletions varied significantly between experiments, consistent with an origin during library construction (Table 2). Consequently, while enough real deletions are present to establish a correlation with UR regions, we sought to identify a more representative and highly enriched collection of deletions associated with UR.

Analyzing rearrangement breakpoints

Identifying junction reads spanning salivary gland-specific deletions appeared to be a way to increase specificity, since DNAs generated by random ligation should be larger on average than individual DNAs and end sequence reads might only rarely be long enough to reach artificial junctions. Furthermore, junction sequences potentially

provide information on the mechanism of break repair. However, most alignment programs such as Eland and BWA do not efficiently identify junction reads. Such reads end up as unaligned or as partially aligned depending on the location of the junction and the parameters of the alignment algorithm.

Empirically, we found that carrying our a BLAT or BLAST searches with unaligned or partially aligned reads frequently revealed new alignment information. Consequently, we scrutinized all high-quality unmapped reads from read-pairs with only 1 aligned read, as well as all reads in which at least 15 bp was unmatched in euchromatic genomic regions (Table 2). By re-aligning these 770,921 reads to the genome using BLAT we identified 3,594 reads from salivary gland DNA that spanned the breakpoint of a deletion 10-500 kb in size with 99-100% sequence matches on both ends. Identical treatment of embryo DNA reads yielded only 388 potential breakpoint reads.

If the salivary gland deletions are generated by incomplete replication and represent the molecular mechanism of sequence under-representation, then they should be preferentially located within the UR regions. Plotting the position of each deletion and comparing their

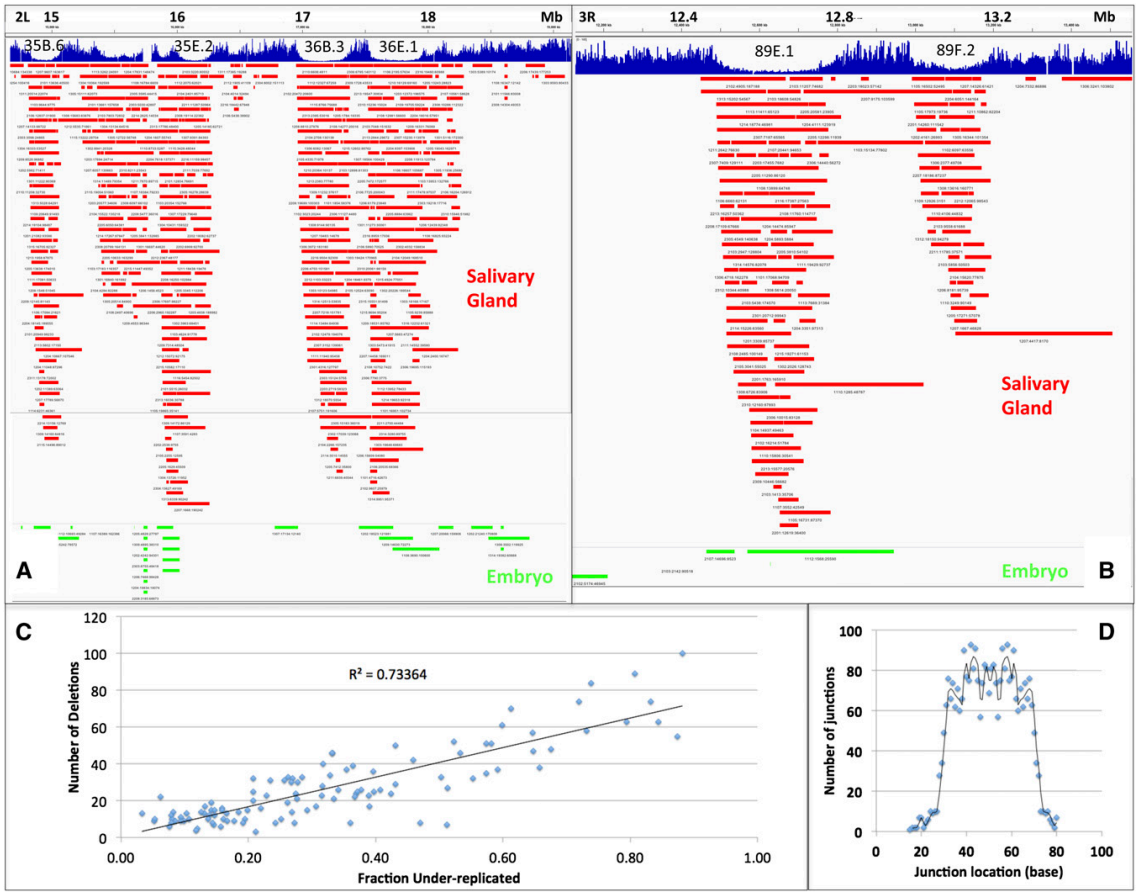


Figure 5. Replication fork breakage and repair extensively alter the genome structure of polytene DNA

A) Chromatin types are shown as in Figure 2C; below a histogram of SG read density in 2L:15-19 Mb shows the

Figure 5 continued from previous page

location of UR regions. Underneath, deletions identified based on the indicated breakpoint read from salivary gland DNA (red) or embryo DNA (green) are shown. Together, the deletions reproduce the under-replication pattern. B) Chromatin organization and under-replication shown as in A, but surrounding UR regions 89D.5 and 89F.2. The specificity of the deletions for UR regions is clear. C)

Figure 5 continued from previous page

The number of salivary gland-specific deletions in UR regions (excluding deep heterochromatin; i.e. states 8 and 9) depends on the degree of under-replication. D). Deletions on the same strand (red) and that switch strands (green) are positioned differently on average in URs such as 87A.5. A diagram at right shows how these two types of deletions may arise from stalled replication forks that break and undergo repair (arrows).

Text continued from page 76

number and location to a salivary gland read profile that highlights the location of UR regions showed that this was indeed the case (Figure 5A, B). For example, in UR-rich region 35 and 36 on chromosome 2L, the distribution of deletions precisely mimics the location of UR junctions (Figure 5A). Most deletions are located within a single UR, but in the case of nearby regions, such as UR36B.3 and UR36E.1, at least 6 deletions that span the two zones are evident (Figure 5A). Equally, precise localization is observed in the Ubx region (UR89E.1) and nearby UR 89F.2 (Figure 5B). In contrast, the Dfs identified from embryo DNA were not enriched in UR regions. Those appearing in small clusters appeared to be alignment artifacts associated with regions containing short tandem repeats separated by 10-500kb, and these deletions were seen in similar numbers in the salivary gland as well.

If we are detecting the deletions that give rise to salivary gland sequence under-representation, then every UR should contain an excess of deletions compared to the embryo controls or to normally replicated regions. We calculated the number of deletions from salivary gland and embryo DNA that lie wholly or partially within each UR region and found this to be the case (Table 4). (The only

exception was UR42B containing the major Drosophila piRNA locus, a highly repetitive region). Moreover, the number of deletions was obviously related to the level of sequence reduction. For example, in major UR regions, such a UR89E.1 containing the bithorax complex, 61 deletions in salivary gland DNA recapitulate the read profile, but only one partially overlapping deletion was found in embryo DNA. The 350 kb just prior to *Ubx*, which does not under-replicate contains no salivary gland deletions (Figure 5B). Similarly, in region 36 of chr2L, there were 55 deletions with UR36B.5 and 65 in UR36C.10 (compared with 2 in embryo DNA), while only 10 Df were found in the first 1 Mb of chr2L which lacks URs. Weak URs contained a correspondingly smaller, but still significant number of deletions and these values were strongly correlated (Figure 5C). Even the 33 weakest URs that could be identified by the ratio method, which had under-replication values less than 1.2, still contained an average of 11.7 deletions in salivary gland, but only 0.63 in embryo. Indeed, we observed additional clusters of 4-8 deletions that may correspond to real UR regions to weak to find by the read ratio method. Overall, 2,811 of the 3,594 salivary gland deletions (78%) were located within the mapped UR regions.

The re-alignment of putative breakpoint reads also identified candidate translocation junctions. However, most of these joints were unlikely to correspond to true translocations generated during salivary gland under-replication in vivo. There was no enrichment for such translocations in the salivary gland since a total of 13,379 candidate junctions were identified within euchromatin in the three salivary gland experiments, while 14,822 candidates were identified in the corresponding embryo DNAs. Moreover the chromosomal location of the putative translocation pairs among the major euchromatic contigs appeared to be random in DNA from both sources (data not shown). We concluded that the great majority of putative translocation junctions were the result of background ligation during library preparation. A low level of real translocations caused by repair of broken forks in UR regions to broken forks originating on other chromosomes would have been hidden by this background.

Properties of UR-associated deletions

Analysis of the salivary gland deletion junctions defined by our breakpoint analysis revealed additional information. When plotted the location of the deletion junction within the 100bp sequence, nearly all the junctions fell between position 30 and 70 (Figure 5D).

Recovery of more asymmetric deletion junction reads using our methods must be much less efficient.

About half of the junctions occur at sites with no nucleotide overlap, while the remaining joints show very limited homology of 1-6 bp and no evidence of a consensus. Only 6% of joints showed homology at the site of joining that was greater than 6 bp. 1,706 of the deletions involved no change of strand orientation and these tended to be larger and to span the edges of the UR (Figure 6A). Such molecules may be generated when two approaching forks on a single DNA molecule stall near each edge of a UR, undergo breakage and the two free ends join to each other. In contrast, 1,887 deletions switched strands. Presumably, when more than one stalled fork is present on the same edge of a UR region, forks may break and resolve by ligating to a free end originating from a different fork on the same side, thereby generating a giant inverted repeat (Figure 6B). Sequences within the UR would be lost as a consequence of such events that are not included between the breakpoints of this class of deletion. This would also rearrange the chromatin and binding sites which would in turn reprogram the DNA affected and the regions flaking the genomic rearrangement.

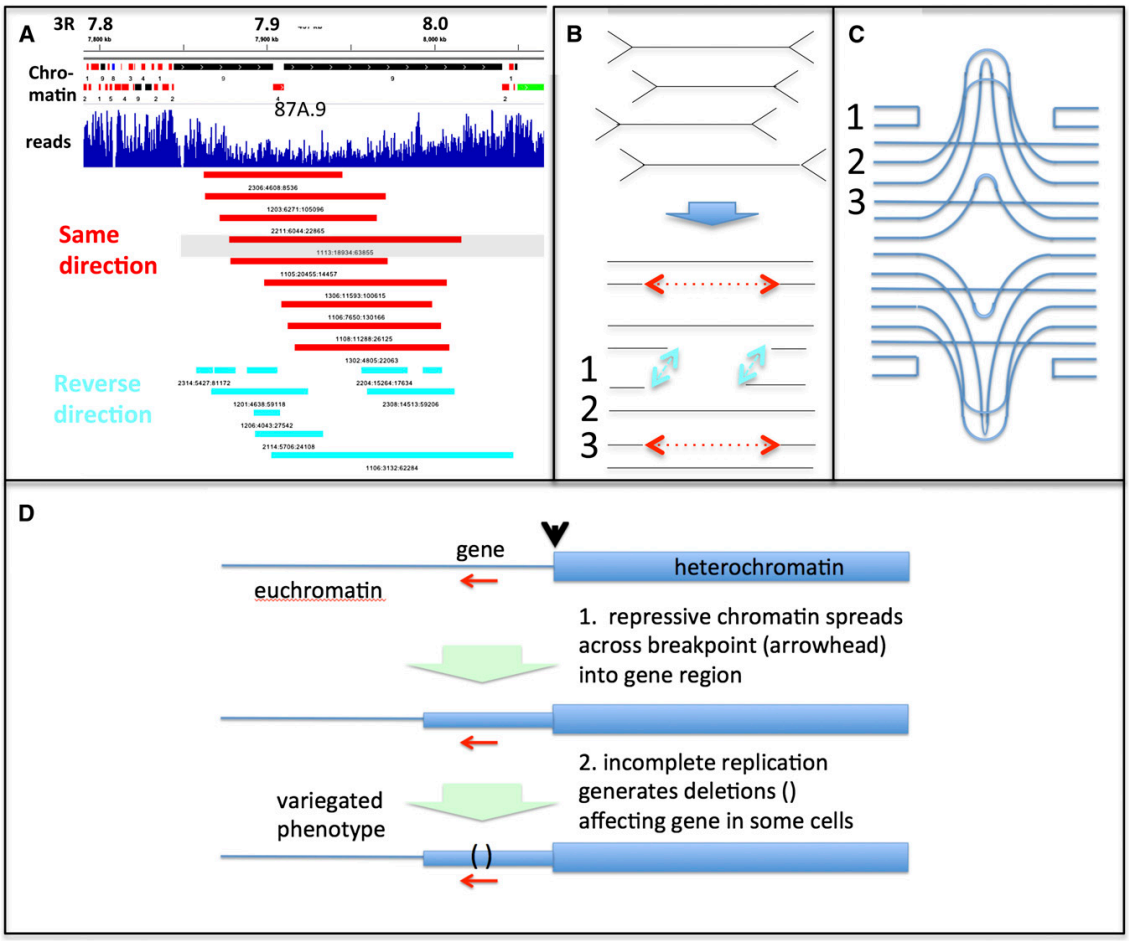


Figure 6. UR-generated deletions alter gene expression

A) The genome region surrounding UR84D.6 is shown; chromatin state and UR regions at the top are as in Fig. 3. Below, gene expression (blue bars) in larval fat body and larval midgut is plotted from (Nordmann et al. 2011). Fat body expression of *Antp* is likely to be affected by UR-generated deletions (red). B) The genome region surrounding UR87A.5 is shown as in A. Fat body expression of the IgSF gene dpr-17 would be disrupted or altered by many of the UR-generated deletions (red). C) The genome region surrounding UR36C.10 is shown as in A. Deletions identified

Figure 6 continued from previous page.

from junction reads (red) illustrate how changes in inhibitory or regulatory sequences within a UR might alter the expression of active genes adjacent to the region. The IgSF gene fasciclin III is shown relative to the many deletions (red) that might affect its expression. Several genes within the UR expressed in larval fat body would also be affected by the deletions. D. Model of position-effect variegation where spreading heterochromatin causes DNA under-replication. Variegated gene expression would result from the somatic deletions generated in the new UR region.

Text continued from page 82

UR zones correspond closely to chromatin states and are enriched in Hox and IgSF genes

The UR regions showed an exceptionally close correlation with chromatin states and could be readily classified on this basis (see Table 4; Methods). A small fraction of URs were associated with Polycomb domains (state 6 of Karchenko et al. 2012), while most pericentric regions were classified as heterochromatic (state 8-9). The majority of the UR regions were comprised of state 9 chromatin, also known as "black" chromatin, which has been associated with reduced average gene density and transcriptional activity, but which is capable of tissue-specific expression (Karchenko et al. 2012; Filion et al. 2011).

The more complete census of salivary gland UR sites made it possible to search the genome more thoroughly than previously possible for classes of genes that are associated with these regions (Table S4). We confirmed that both the major Drosophila Hox gene clusters, the Antennapedia Complex and the Bithorax Complex, which reside in the genome's largest Polycomb domains, are UR regions, as well as 13 additional Polycomb domains encoding genes such as *disco, disco-r, upd, upd2, H15, CG31647, CG14926,*

salm, salr, vvl, lov, knrl, kni, tsh, CG16778, and *Ptxl*. However, many other Polycomb domains did not under-replicate at detectable levels. Of particular note, IgG superfamily (IgSF) proteins (Ozkan et al. 2013) are strongly enriched in UR regions. In particular, among the four classes of IgSF proteins characterized by Ozkan et al (2013), 13 of 14 Beat-class, 7 of 8 Side-class, 5 of 11 DIP-class, and 9 of 20 Dpr-class IgSF genes are in UR regions (Table S4). Since UR regions constitute only 18.9% of euchromatin, finding 34 of 53 IgSF genes in UR regions is unlikely to occur by chance ($p<10E-13$; Binomial distribution). Other genes involved in cellular adhesion including proteins involved in IgSF and LRR interactions such as Robo, Robo3, Lea, and Caps, as well as CadN, CadN2, Connectin, Fred, Ed, Rst, Kirre and Snap25) are also encoded within UR regions. Many genes in UR regions, including Hox genes and IgSF genes, are unusually large due to the presence of large introns.

DISCUSSION

Polytene salivary gland cells undergo extensive genomic alterations during development

Our results show that larval salivary gland cells covalently alter their somatic genome structure at hundreds of sites within 112 dispersed euchromatic domains (Table

4). The affected zones, known as under-replicated regions, were first identified because they display reduced copy number relative to the ploidy of the genome as a whole (Gall et al. 1971; early Zhimulev work; etc. reviewed in Spradling and Orr-Weaver, 1987). By deep sequencing we report a more complete census of UR regions, revealing that they constitute 19% of euchromatin and house at least 10% of Drosophila genes. In striking contrast to the assumption that under-replicated regions contain only stalled, nested replication forks ( ), we find that all 112 UR regions contain a diverse array of DNA deletions at levels sufficient to entirely explain the copy number changes. Similar deletions are seen only at much lower levels, if at all, in UR region DNA from early embryos or at non-UR euchromatic sites within the salivary gland.

The sufficiency of the observed deletions to explain the copy number reductions is based on simple calculations. From three salivary gland sequencing experiments with a total read depth of 239, we observed 3,659 deletions defined by junction reads, of which 2,811 overlapped defined UR regions. However, the distribution of beakpoints within these reads (Figure 5D) showed that only breaks within the central 40% of the read had been efficiently recovered. So a better estimate of the total

number of deletions in UR regions would be 2,811/0.4 = 7,027. For comparision, if all copy number variation in the UR regions results from deletions, then 8,861 deletions should be observed in the pooled sequences from our three salivary gland DNA sequencing experiments, given their read depth (see Methods). The close agreement of these numbers indicates that the great majority of sequence under-representation does result from deletions, rather than from nested forks, or free DNA ends (neither of which will generate novel junctions upon sequencing).

Stalled replication and repair generates UR-associated deletions

A great deal of evidence had accumulated previously that polytene under-replication is caused in some way by replication fork stalling. The most strongly affected regions replicate late in S phase (Belyaeva et al. 2008), which renders them susceptible to remaining unfinished when the endocycle S phase is completed. Furthermore, UR zones are characterized by repressive chromatin (Sher et al. 2012;), a property we confirmed for all but one of the 112 UR regions (Table 4). Origins are sparse and replication forks may have difficulty elongating through these regions in polytene cells (Sher et al. 2012). We observed that the domain boundaries of many strong UR regions corresponds

closely to the junctions of dozens of deletions, suggesting that replication forks frequently stall almost immediately after encountering a UR domain. UR regions are also known to be sites of elevated DNA repair. Salivary gland chromosomes contain elevated amounts of phosphorylated His2Av, and sites enriched in phosphorylated His2Av correspond to many UR regions (Andreyeva et al. 2007). However, earlier studies could not determine whether the repair activity maintained nested fork structures or if some broken DNA ends persisted without completing repair.

Our experiments demonstrate that DNA replication forks are unstable in polytene cells, but can be efficiently repaired. We observed high levels of fork instability during chorion gene amplification that is followed by repair to form deletions of heterogeneous size. Many of the deletions begin just a few kb from the site of amplification initiation near Cp18. Overall, about 2% of all amplified DNA strands contain a deletion in our experiments. Repair was found to take place efficiently within a relatively short distance from broken ends. In UR regions, damage repair by end joining occurs predominantly within the same UR. The efficient breakage and repair of stalled forks probably explains why nested replication forks from amplified chorion genes were readily observed by

electron microscopy of ovarian follicle DNA (Osheim and Miller, 1983) but were never reported from salivary gland DNA, either using electron microscopy or by 2D-gel analysis of a UR region (Glaser et al. 1992).

Other studies prior to the advent of high throughput sequencing provide further support that DNA under-replication causes DNA alterations. Given the observed heterogeneity of UR-generated deletions and their dispersal across large local regions, it is not surprising that small restriction fragments undergo detectable changes in abundance, but not in size (Spierer and Spierer, 1984; Karpen and Spradling, 1990). However, by including a UR region on a minichromosome, or by analyzing large restriction fragments within heterochromatic regions, it was possible to observe consistent DNA changes in polytene DNA, that generated heterogeneous-sized molecules (Spradling, 1993; Leach et al. 2000; Andreyeva et al. 2008).

DNA alterations are likely a characteristic of most polyploid Drosophila cells

Recent studies of DNA replication shows that both initiation sites and replication timing is quite similar in 3 cell lines ( ), and in salivary gland cells (orr weaver). ; although small differences in UR were

reported between tissues (Sher et al) these studies used a cutoff of 2x for underreplication. Somatic DNA alterations are not confined to salivary gland cells, but probably occur in all polytene cells, which are thought to include all differentiated larval as well as many adult cells, including salivary gland, follicle cell, fat body, prothoracic gland, midgut, pericardia, rectal gland, stomach, gastric caeca, and hindgut (reviewed in Ashburner, 1970). Previous studies have directly documented under-replication of satellite DNA, rDNA and other sequences in several such cells and they display typical polytene chromosomes (reviewed in Orr-Weaver and Spradling, 1989). The one exception is early nurse cells, which prior to endocycle 5 replicate fully and lack typical polytene chromosomes (Dej and Spradling, 1999). More recent studies reveal that many of these cells as well as diploid tissue culture cells unexpectedly share a similar distribution of replication origins (Nordmann et al. 2011; Sher et al. 2012), late replication (Eaton et al. 2011), and heterochromatin (Karchenko et al. 2012). Despite the larger number of UR regions identified here, we confirmed that a repressive chromatin state in nearly all of them correlates strongly (and likely controls) under-replication in salivary gland cells. Mostly the same regions replicate

late in S phase, and under-replicate in polytene cells (Nordman et al. 2011; Sher et al. 2012). In fact, our more sensitive data show that the even the tissue-specific URs reported by Nordman et al. (2011) are all under-replicated at some level in salivary gland. This suggests that the program of late replication and under-replication is similar in most polyploid cells, but the frequency of fork arrest may be modulated quantitatively by tissue-specific factors such as chromatin state. We also directly showed that polytene DNA from late stage ovarian follicles under-replicates many of the same strong UR zones as salivary gland cells and also contains somatically generated deletions. Thus, the polyploid cells that make up a significant fraction of most Drosophila tissues undergo somatic DNA changes generated in a locally variable but globally directed manner by incomplete replication.

Our studies show that about 24 Mb of the 119Mb within the main chromosome contigs, most of it euchromatic, under-replicates and generates somatic DNA alterations in salivary gland, ovary and probably most other polyploid cells. Even greater levels of under-replication take place in heterochromatic, peri-centromeric regions, where some sequences fail to duplicate in half of all endocycles. While it has not been possible to map the structural

alterations in these regions precisely, due to their repetitive nature, its seems likely that DNA alterations would be generated even more frequently. Since, these regions contain many more sequences that are shared between chromosomes, translocations between chromosomes might be much more common than in the case of unique sequence euchromatin.

Stochasticity of under-replication enhances genetic diversity within individual cells

The similarity in under-replication within both 2nd and 3rd instar salivary glands except in the most strongly under-replicated regions suggests that under-replication occurs extensively before the second instar, and probably takes place throughout all the salivary gland endocycles. Indeed, a failure to undergo a full doubling of DNA content during the first endocycle has been documented in many types of polyploid cells (reviewed in Spradling and Orr-Weaver, 1987). This suggests that each UR has a low intrinsic probability of incomplete replication during every endocycle, and that UR values simply represent the effects of their propensity averaged over many cells. This simple situation will result in a wide range of deletion abundances within individual polyploid cells and that are likely to greatly enhance their significance.

Consider, a UR that fails to replicate and generates a deletion on average only once every 10 endocycles. In a population of ten ideal cells, one cell will gain a deletion on the first endocycle, and upon subsequent replication it will constitute 50% of its DNA at L3. Another cell will generate a different deletion on the 2nd endocycle, which will occupy 25% of its strands. A third cell will generate a deletion on the third endocycle, etc. The net result is that all ten cells will contain a single unique deletion, but their abundance will vary from 0.2% to 50%. The genetic changes in at least several of the cells are potentially very significant, but such a UR would be scored as only having an under-replication value of 1/0.9 = 1.1. If the cell has twice that tendency to stall, and for simplicity we again assume that only one deletion occurs per cell, then two cells would have deletions at 50%, two at 25%, etc. Now there are twice as many strongly affect cells, yet the average under-replication value of the region would have only increased to about 1/0.8 = 1.25. In fact, in order to generate high underreplication values, it is necessary that multiple deletions occur in the same cell, i.e. such that the wild type strands surviving the first endocycle intact will frequently fail to complete replication in subsequent rounds. Thus, in UR regions with

high under-replication values, most cells contain multiple rearrangements at quite high copy number, with others at diminishing levels.

These considerations show two important points. First, even loci with very low average UR values will generate potentially significant genetic diversity, and in some cells those novel products will be present at high copy number. Second, these same calculations apply to each of the 122 URs independently. Given, the presence of so many UR regions, and the fact that more than 25 have very high values, i.e. more than 2, it follows that every cell will have dozens of unique deletions at high copy number, and many times more at lower copy number. Previously, regions with low under-replication values were thought to be unimportant, because the only effects of under-replication was assumed to be on gene dosage. Now that we know under-replication generates genomic novelty, and that stochasticty spreads that novelty among at least some cells at high levels, the true potential of under-replication to generate significant variation can be seen to be extraordinary. This expected cell to cell variation is not simply theoretical, but has been observed. The level of under-replication of particular classes of rDNA repeat varied stochastically from cell to cell in the salivary

gland (Belikoff and Beckingham, 1985). The copy number of the yellow gene when located in a UR region varied widely between individual salivary gland cells when assayed by in situ hybridization, and yellow expression variegated in polytene bristle cells (Karpen and Spradling, 1990).

Significance for polytene chromosome structure

Our finding that tens of thousands of new junctions are produced in each polytene cell strongly supports the idea that somatic DNA modifications underlie polytene morphology (Ashburner et al. 1972; Spradling et al.1992; Leach et al. 2000; Andreyeva et al. 2008). Cytogenetic regions 35 and 36 on chr2L were found to contain the most frequent and deepest UR regions, and to contain many deletions within and between nearby URs. These polytene regions are most subject to structural disruption and ectopic fiber formation in salivary gland chromosomes (Figure 1A), suggesting that ectopic fibers result from the mis-pairing caused by strands of very different length and sequence content. Further improvements in methods, such as the application of long-range sequence reads, might allow more detailed mapping of the changes going on in highly repetitive regions that probably underlie chromocenter formation.

A new model of position effect variegation

When euchromatic genes are rearranged near heterochromatin, they usually display the phenomenon of position-effect variegation (PEV). In recent decades, only epigenetic explanations for PEV have been considered (Ahmad and Golic, 1996; reviewed in Elgin and Reuter, 2013). However, previous evidence suggested that somatic genetic changes related to under-replication contribute to some examples of PEV, for example when the affected gene is located in a new UR generated by the rearrangement (Karpen and Spradling, 1990; Glaser et al. 1992; Spradling, 1993; Glaser et al. 1997; Leach et al. 2000). *SuUR* mutation suppresses both the under-replication and PEV of several rearranged genes (Belyaeva et al. 2008). We propose a model for PEV that incorporates both the current postulate of chromatin spreading, as well the phenomena of somatic mutation described here. Near the breakpoint of a chromosome rearrangement juxtaposing heterochromatin and euchromatin, chromatin spreading would invade adjacent euchromatin, as currently envisaged (Figure 6D). However, we propose that the reason the effects are variegated rather than uniform is because the new domain of repressive chromatin causes replication forks to arrest and undergo

repair into somatic DNA alterations that differ between cells.

Somatic DNA instability may be functional

Under-replicated regions have been likened to peri-centromeric heterochromatin (Belyaeva et al. 2008). They replicate late in S phase (Zhimulev et al. 1982; contain many transcriptionally silent genes (Belyakin et al. 2005; Sher et al. 2011) and their chromatin is enriched in repressive histone modifications (Belyaeva et al. 2008; Sher et al. 2011; Maksimov et al. 2012). It has remained a matter of debate whether under-replication affects genes that are actually expressed in under-replicated cells. Our studies provide new insight into this question by precisely defining both the chromosome regions that are altered by under-replication and the molecular nature of the changes, which are much more complex than simple copy number reductions of otherwise unaltered genes. Each deletion reduces the expression of the genes that were lost, and potentially alters the regulation of multiple genes located near the breakpoints. Indeed, our data strongly argues that under-replication does affect the structure and activity of hundreds and probably thousands of genes in diverse polytene tissues (Figure 6).

The non-random association of IgSF genes with UR zones suggests that somatic DNA changes may be used during development to increase genetic variation. IgSF genes are expressed on cell surfaces and govern cellular interactions that are important during multiple developmental processes, especially in the nervous system (Ozkan et al. 2013). Beat proteins are expressed in different subgroups of neural cells (Pipes et al. 2001) and some individual family members have been shown to guide motor axons (Fambrough and Goodman, 1996). Combinations of IgSF molecules may affect synaptic adhesion (Yamagata et al. 2003). Although thousands of hemophilic and heterophilic binary combinations of IgSF proteins are possible based on the genes encoded in the germline genome, additional diversity is known to be generated in these protein families in the vertebrate immune system (Alt et al. 2013). In Drosophila, the Dscam family of IgSF proteins is extensively diversified by a different mechanism- differential splicing (Wojtowicz et al. 2007). Consequently, a potential reason for locating such genes in UR regions would be to use incomplete replication and repair to generate somatic diversity of IgSF protein structure or expression within the individual cells of one or more tissues. The low concentration of any particular alteration suggests that

the production of gain-of-function changes would be favored, including the production of a protein in a novel location, or with a unique property.

The gene structure of IgSF genes and other adhesion protein genes in UR regions might lend itself to such a purpose. Many of these genes, as well as other genes located in URs such as HOX genes, are characterized by long transcription units with multiple long introns. Large genes would provide a greater cross section for regionally localized rearrangements to generate fusion genes with altered expression patterns and/or coding capacity. Gene organization, chromatin structure and replication timing may have been optimized by evolution to generate potentially useful diversity using a semi-random mechanism. Because the replication program appears to be similar between cell types, genes might undergo similar rearrangements in multiple tissues, including many where they are inactive; only in a critical tissue would they be both rearranged and expressed. Most Drosophila neurons are thought to be diploid, although highly polyploid neurons are known to exist in organisms such as *Aplysia californica*. However, polyploid cells are common in the Drosophila brain. One potential source is glial cells, which can be polyploid (Unhavaithaya and Orr-Weaver, 2012).

Surface protein or other genetic diversity in glia might affect neural development and organization.

Under-replication may be conserved in evolution

After this manuscript was submitted for publication, Baker et al. (2014) reported that polyploidy mouse giant trophoblast cells from the placenta contain at least 47 under-replicated regions dispersed throughout most of the chromosomes.

Recently, evidence has been found that replication timing, including late replication, strongly affect the accumulation of mutations in diploid cells (reviewed in Lambert and Carr, 2012). Comparisons between closely related Drosophila species show that late-replicating regions have greater sequence diversity (Weber et al. 2011). In cancer cells, some of which lack P53 function like the Drosophila salivary gland (Mehrotra et al. 2008), large-scale structural variations are greatly increased in late-replicating genomic regions and preferentially join regions with proximity in the nucleus (De and Michor, 2011). The diverse distribution of DNA breaks at the sequence level in well-studied cancer rearrangements (Ross et al. 2013), can be explained if breaks are initially generated by replication fork arrest, as in the diverse breakpoints observed following salivary gland

underreplication. Consequently, it seems likely that diploid somatic and germline cells in rare circumstances follow the same pathway as polytene salivary gland cells due to replication fork stalling, breakage and error-prone repair. Thus, by studying polytene chromosomes, which are observed in only a small number of species distant from mammals, one can learn about a process crucial to carcinogenesis and genome evolution.

MATERIALS AND METHODS

Drosophila Strains

Strain iso-1: *y[1]; Gr22b[1] Gr22d[1] cn[1] CG33964[R4.2] bw[1] sp[1]; LysC[1] MstProx[1] GstD5[1] Rh6[1],* *y w: y[1]w[1],* *y ry: y[1]; ry[506]*, and *SuUR*$^{-}$: *In(1)scV2, scV2*; *SuUR*^{ES} were obtained from the Bloomington stock center. Stocks were maintained on standard fly food, which was supplemented with additional yeast beginning one week before tissue collection.

DNA Isolation

All samples was prepared independently. Embryos were collected using grape juice agar caps during a 0:30- 2:30 collection window, and dechorionated using bleach. Salivary glands were dissected from 300-400 larvae per preparation (late third instar except as noted) in cold Grace's Insect Media (Life Technologies). Adult females were anesthetized using CO2 and ovaries dissected and late stage follicle separated using jewler's forcepts.. Tissues were flash frozen in liquid nitrogen immediately after collection. DNA was purified using a QIAamp DNA Mini Kit (Qiagen) and RNAse A (Qiagen) according to the manufacturer's protocol.

DNA and library preparation: Libraries for paired end sequencing were prepared using Illumina's TruSeqDNA Sample

Prep Kit LT using the LS and the gel-free options. Fragmentation was carried out with a Diagenode Bioruptor sonicator, with power setting "Low", at 4 °C, for 15 minutes (intervals of 30 seconds of sonication, 30 seconds without sonication over 30 minutes). 100ng of sonicated DNA was used as input. The amplification reaction was altered (while still using kit reagents and cycling conditions) to the following: (ligated DNA 3ml, PCR Master Mix 25ul, PCR primer cocktail 5ul, Resuspension Buffer 17 ml). The size of the resulting DNA was 300-500 bp, including 120 bp of linkers. Most libraries were prepared one different days, but RefSG2, RefSG3, RefEmb2 and RefEmb3 were prepared in the same batch.

Chromatin domain analysis:

The chromatin state of UR and surrounding regions was classified based on the 9-state model described in Kharchenko et al. (2011) and applied to the S2 cell line. For further simplicity in viewing, similar states were grouped as follows: red (“active:” states 1-5); green (“Polycomb regulated:” state 6); blue (“heterochromatic:” states 7-8); black (state 9). Note that state 9 corresponds to the “black” chromatin described by Filion et al (2010). When downloaded chromatin data was displayed relative to salivary gland read data, there was a striking

correspondence to chromatin domains (Figure 3). The predominant chromatin state associated with each UR is listed in Table 4.

Alignment and identification of UR regions:

Sequencing was carried out on an Illumina HiSeq2000 and reads were aligned to the Drosophila genome version R5.33 using ELAND or BWA software and visualized using IGV. UR regions were automatically flagged if the salivary gland /embryo read average fell more than 2 standard deviations from the mean for 5 or more consecutive 5kb windows. Most such regions have a characteristic shape profile, consisting of a monotonic decrease to a minimum value near the center, followed by a similar rise back to the baseline (Fig. 1E-G). In addition, less than 10 of the weakest UR regions were added because the shape of their deviation from baseline over 70-200 kb appeared very similar to those of stronger UR regions. Overlapping URs were separated by at the local maximum point (Figure 1E). The depth of each UR region was determined as the ratio of its low point and baseline. Under-replication was the inverse of this value. Further analyses were carried out using Samtools, IGV tools, FileMaker Pro, Microsoft Excel, PYTHON, C++ , and the Unix command line (see below).

Prediction of deletion number: For a UR region that is 2X under-replicated, half the DNA strands must contain a deletion (or other rearrangement). Hence the minimum number of predicted deletions would be 0.5 times the read depth (there could be additional smaller deletions off center). For an under-replication value of n, the corresponding formula for the fraction of strands with deletions is f =1-1/n. We used the under-replication values in Table 4 to calculate f for each UR, and multiplied by the mean single copy read depth (after correction for PCR duplicates) to yield the predicted number of unique deletions (Table 2).

Analysis of deletions defined by anomalous reads: Read pairs aligned but separated by 10kb-500kb were sorted by AWK and loaded into a custom FileMaker Pro database with all associated sam data. Reads pairs aligned to heterochromatic regions or mitochondrial DNA were removed, as were reads with quality fields containing more than 4 #'s, and duplicate read pairs. Both reads matched the genome 100% in more than 95% of these pairs. Scripts were used to output display files. Each deletion is denoted by the shortened name of the read pair from which it is derived (i.e. HWI-ST375:119:D0A2LACXX:6:2205:1397:93872 is shown as 2205:1397:93872.)

Identification of breakpoint reads: For each experiment, all reads from pairs mapped within euchromatin that were unaligned, and all mapped euchromatic reads in which a run of 15-80 bases were not aligned ("soft-clipped") were collected using AWK, and the 100bp sequences were aligned to the Drosophila 5.33 genome locally using BLAT. The BLAT output was parsed at six matches per sequence and loaded along with the corresponding sam file data into a FileMaker Pro database. Deletions identified by BLAT were accepted if Tgapbases was between 10k and 500k, and if one segment of the deletion-containing read was on the same arm and at a proper distance from the mate. Additional deletions were identified if BLAT aligned the sequence on two separate lines corresponding to genome regions on the same arm separated by between 10kb and 500kb, and that together matched at least 97% of the input sequence. One of the matching segments was required to be on the same arm and within a proper distance from the mate. More than 95% of the deletions matched 100% of the input sequence and 100% of the target, with the deletion as the only gap. However, candidate deletions involving sequences with more than BLAT 4 matches or involving input sequences with long homopolymeric tracts were excluded. Scripts were used to output display files.

REFERENCES

Aggarwal B.D. and Calvi B.R. (2004) Chromatin regulates origin activity in Drosophila follicle cells. Nature 430: 372–376.

Ahmad K. and Golic K. 1996. Somatic reversion of chromosomal position effects in Drosophila melanogaster. Genetics 144: 657-670.

Alt F.W., Zhang Y., Meng F.L., Guo C., Schwer B. 2013. Mechanisms of programmed DNA lesions and genomic instability in the immune system. Cell 152:417-29.

Andreyeva E.N., Kolesnikova T.D., Belyaeva E.S., Glaser R.L., Zhimulev I.F. 2008. Local DNA underreplication correlates with accumulation of phosphorylated H2Av in the Drosophila melanogaster polytene chromosomes. Chromosome Res. 16: 851-62.

Belikoff E.J. and Beckingham K. 1985. A stochastic mechanism controls the relative replication of equally competent ribosomal RNA gene sets in individual dipteran polyploid nuclei. Proc Natl Acad Sci U S A. 82: 5045-9.

Belyaeva E.S., Zhimulev I.F., Volkova E.I., Alekseyenko A.A., Moshkin Y.M., Koryakov D.E. 1998. Su(UR)ES: A gene suppressing DNA underreplication in intercalary and pericentric heterochromatin of Drosophila melanogaster polytene chromosomes. Proc. Natl. Acad. Sci. USA 95: 7532–7537.

Belyaeva E.S., Boldyreva L.V., Volkova E.I., Nanayev R.A., Alekseyenko A.A., Zhimulev I.F. 2006a. Effect of the Suppressor of Underreplication (SuUR) gene on position-effect variegaion silencing in Drosophila melanogaster. Genetics 165: 1209-20.

Belyaeva E.S., Demakov S.A., Pokholkova G.V., Alekseyenko A.A., Kolesnikova T.D., Zhimulev I.F. (2006) DNA underreplication in intercalary heterochromatin regions in polytene chromosomes of Drosophila melanogaster correlates with the formation of partial chromosomal aberrations and ectopic pairing. Chromosoma 115: 355-366.

Belyaeva E.S., Andreyeva E.N., Belyakin S.N., Volkova E.I., Zhimiulev, I.F. 2008. Intercalary heterochromatin in polytene chromosomes. Chromosoma 117: 411-418.

Belyakin S.N., Christophides G.K., Alekseyenko A.A., Kriventseva E.V., Belyaeva E.S., et al. (2005) Genomic analysis of Drosophila chromosome underreplication reveals a link between replication control and transcriptional territories. Proc Natl Acad Sci U S A 102: 8269–8274.

Calvi B.R. and Spradling A.C. (1999). Chorion gene amplification in Drosophila: A model for metazoan origins of DNA replication and S-phase control. Methods. 18: 407-17.

Chalker D.L. and Yao M.C. (2011) DNA elimination in ciliates: transposon domestication and genome surveillance. Annu Rev Genet. 45: 227-46.

Claycomb J.M., MacAlpine D.M., Evans J.G., Bell S.P., Orr-Weaver T.L. (2002). Visualization of replication initiation and elongation in Drosophila. J Cell Biol. 159: 225-36.

De S. and Michor F.: DNA replication timing and long-range DNA interactions predict mutational landscapes of cancer genomes. Nat Biotechnol 2011, 29:1103-1108.

Duncan A.W., Hanlon Newell A.E., Smith L., Wilson E.M., Olson S.B., Thayer M.J., Strom S.C., Grompe M. 2012. Frequent aneuploidy among normal human hepatocytes. Gastroenterology 142: 25-8.

Elgin S.C. and Reuter G. 2013. Position-effect variegation, heterochromatin formation and gene silencing in Drosophila. Cold Spring Harbor Perspect. Biol. 5: a017780.

Filion G.J., van Bemmel J.G., Braunschweig U., Talhout W., Kind J., Ward L.D., Brugman W., de Castro I.J., Kerkhoven R.M.,

Bussemaker H.J. and van Steensel B. 2010. Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell 143: 212–224.

Fox D., Gall J.G., Spradling A.C. 2010. Error-prone polyploid mitosis during normal Drosophila development. Genes Dev. 24: 2294-2302.

Gall J.G., Cohen E.H., Polan M.L. (1971) "Repetitive DNA sequences in Drosophila" Chromosoma 33: 319- 344.

Gilbert S. 2013. Developmental Biology. 10th ed. (Sinauer Associates, Inc.: Sunderland, MA). 750pp.

Glaser R.L., Karpen G.H., Spradling, A.C. 1992. Replication forks are not found in a Drosophila mini-chromosome demonstrating a gradient of polytenization. Chromosoma 102: 15-19.

Glaser R.L., Leach T.J., Ostrowski S.E. 1997. The structure of heterochromatic DNA is altered in polyploid cells. Mol. Cell Biol. 17: 1254-63.

Grandi F.C. and An W. 2013. Non-LTR retrotransposons and microsatellitels: partners in genomic variation. Mob. Genet. Elements 3: e25674.

Greshn A.F., Prescott D.M., Oka Y., Loukin S.H., Chappell J.C. 1989. Reordering of nine exons is necessary to form a functional actin gene in Oxytrichanova. Proc. Natl. Acad. Sci, USA. 86: 6264.

Hammond M.P. and Laird C.D. 1985. Chromosome structure and DNA replication in nurse cells and follicle cells of Drosophila melanogaster. Chromosoma 91: 267-78.

Hunkapiller N.M. and Fisher S.J. 2008. Placental remodeling of the uterine vasculature. Methods Enzymol. 445: 281-302.

Karchenko P.V., Aleksehyenko A.A., Schwartz Y.B., Minoda A., Riddle N.C., Ernst J., Sabo P.J., Larschan E., Gorchakob A.A., Gu T., Linter-Basso D., Plachetka A., Shanower G., Tolstorukov M.Y., Luquette L.J., Xi R., Jung Y.L., Park R.W., Bishop E.P., Canfield T.K., Sandstrom R., Thurman R.E., MacAlpine D.M., Stamatoyannopoulos J.A., Kellis M., Elgin S.C.R., Kuroda M.I., Pirrotta V., Karpen G.H., Park P.J. 2012. Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature 471: 480-485.

Karpen G.H. and Spradling A.C. 1990. Reduced DNA polytenization of a minichromosome region undergoing position-effect variegation in Drosophila. Cell 63: 97.

Kazazian H.H., 2011. Mobile DNA transposition in somatic cells. BMC Biol. 9: 62. doi: 10.1186/1741-7007-9-62.

Kolesnikova TD, Posukh OV, Andreyeva EN, Bebyakina DS, Ivankin A.V. and Zhimulev I.F. (2013) Drosophila SUUR

protein associates with PCNA and binds chromatin in a cell cycle-dependent manner. Chromosoma 122: 55—66.

Laird C.D. 1980. Structural paradox of polytene chromosomes. Cell 22: 869-74.

Lambert S. and Carr A.M. 2012. Replication stress and genome rearrangements: lessons from yeast models. Curr. Opin. Genet. Dev. 23: 132-39.

Leach T.J., Chotkowski H.L., Wotring M.G., Dilwith R.L., Glaser R.L. 2000. Replication of heterochromatin and structure of polytene chromosomes. Mol Cell Biol. 20: 6308-6316.

MacAlpine D.M., Rodriguez H.K., Bell S.P. 2004. Coordination of replication and transcription along a Drosophila chromosome. Genes Dev. 18: 3094— 3105.

MacAlpine H.K., Gordan, R., Powell, S.K., Hartemink A.J., MacAlpine D.M. 2010. Drosophila ORC localizes to open chromatin and marks sites of cohesin complex loading. Genome Res 20: 201-211.

Macosko and McCarroll (2013) Science 342, 564-565

Maksimov D.A., Koryakov D.E., Belyakin S.N. 2013. Developmental variation of the SUUR protein binding correlates with gene regulation and specific chromatin types in D. melanogaster. Chromosoma doi 10.1007/s00412-013-0445-6

Makunin I.V., Volkova E.I., Belyaeva E.S., Nabirochkina E.N., Pirrotta V., Zhimulev I.F. 2002. The Drosophila suppressor of underreplication protein binds to late-replicating regions of polytene chromosomes. Genetics 160:1023-1034.

Marchetti M., Fanti L., Berloco M., Pimpinelli S. 2003. Differential expressionof the Drosophila BX-C in polytene chromosomes in cells of larval fat bodies: A cytological approach to identifying in vivo targets of the homeotic Ubx, Abd-A and Abd-B proteins. Development 130: 3683-3689.

McConnell M.J., Lindberg M.R., Brennand K.J., Piper J.C., Voet T., Cowing-Zitron C., Shumilina S., Lasken R.S., Vermeesch J.R., Hall I.M., Gage F.H. 2013. Mosaic copy number variation in human neurons. Science 342: 632-7.

Mehrotra S., Maqbool S.B., Kolpakas A., Murnen K., Calvi B.R. 2008. Endocycling cells do not apoptose in response to DNA rereplication genotoxic stress. Genes Dev. 22: 3158-71.

Nordman J. and Orr-Weaver T.L. 2011. Regulation of DNA replication during development. Development 139: 455-64.

Perrat P.N., DasGupta S., Wang J., Theurkauf W., Weng Z., Rosbash M., Waddell S. 2013. Transposition-driven genomic heterogeneity in the Drosophila brain. Science 340: 91-5.

Pindyurin A.V., Moorman C., de Wit E., Belyakin S.N., Belyaevab E.S., Christophides G.K., Kafatos F.C., van Steensel B., Zhimulev I.F. 2007. SUUR joins separate subsets of PcG, HP1 and B-type lamin targets in Drosophila. J. Cell Sci. 120: 2344–2351.

Pipes G.C., Lin Q., Riley S.E., Goodman C.S. 2001. The Beat generation: a multigene family encoding IgSF proteins related to the Beat axon guidance molecule in Drosophila. Development 128: 4545–4552.

Reilly M.T., Faulkner G.J., Dubnau J., Ponomarev I., Gage F.H. 2013. The role of transposable elements in health and diseases of the central nervous system. J Neurosci. 33: 17577-86.

Reizel Y., Itzkovitz S., Adar R., Elbaz J., Jinich A., ChapalIlani N., Maruvka Y.E., Nevo N., Marx Z., Horovitz I., Wasserstrom A., Mayo A., Shur I., Benayahi D., Skorecki K., Segal E., Dekel N., Shapiro E. (2012) Cell lineage analysis of the mammalian female germline. PLoS Genet. 8: e1002477.

Ross D.M., O'Hely M., Bartley P.A., Dang P., Score J., Goyne J.M., Sobrinho-Simoes M., Cross N.C., Melo J.V., Speed T.P., Hughes T.P., Morley A.A. (2013). Distribution of genomic breakpoints in chronic myeloid leukemia: analysis of 308 patients. Leukemia 27: 2105-7.

Rudkin, G. (1969). Non-replicating DNA in Drosophila. Genetics 61: Suppl. 227-238.

Schubeler D., Scalzo D., Kooperberg C., van Steensel B., Delrow J., Groudine M. (2002) Genome-wide DNA replication profile for Drosophila melanogaster: a link between

transcription and replication timing. Nat Genet. 32: 438–442.

Sher N., Bell G.W., Li S., Nordman J., Eng T., Eaton M.L., Macalpine D.M., Orr-Weaver T.L. 2012. Developmental control of gene copy number by repression of replication initiation and fork progression. Genome Res 22: 64–75

Siebert M., Banovic D., Goellner B., Aberle H. 2009. Drosophila motor axons recognize and follow a Sidestep-labeled substrate pathway to reach their target fields. Genes Dev. 23: 1052-1062.

Sima J. and Gilbert D.M.. 2014. Complex correlations: replication timing and mutational landscapes during cancer and genome evolution. Curr. Opinion Genet. Dev. 25: 93-100.

Smith C.D., Shu S., Mungall C.J., Karpen G.H. (2007). The Release 5.1 annotation of Drosophila melanogaster heterochromatin. Science 316: 1586–1591.

Spear B.B. and Gall J.G. 1973. In- dependent control of ribosomal gene replication in polytene chromosomes of

Drosophila melanogaster. Proc. Natl . Acad. Sci. USA 70:1359-63.

Spierer A. and Spierer P. 1984. Similar level of polyteny in bands and interbands of Drosophila giant chromosomes. Nature 307:176-78.

Spradling A.C. and Orr-Weaver T. 1987. Regulation of DNA Replication during Drosophila development. Ann. Rev. Genet. 21: 373-403.

Spradling A.C., Karpen G.H., Glaser R., Zhang P. 1992. Evolutionary conservation of developmental mechanisms: DNA elimination in Drosophila. Symp. Soc. Dev. Biol. 50: 39-48.

Spradling A.C. 1993. Position-effect variegation and genomic instability. Cold Spring Harbor Symp. Quant. Biol. 58: 585-596.

Unhavaithaya Y. and Orr-Weaver T.L. 2012. Polyploidization of glia in neural development links tissue growth to blood-brain barrier integrity. Genes Dev. 26: 31-6.

Usov, K.E., Shelkovnikova, T.A., Stegnii V.N. 2011. Analysis of the chromocenter DNA composition in polytene chromosomes of Drosophila orena ovarnian nurse cells. Genetika. 47:475-483.

Weber C.C., Pink C.J., Hurst L.D. 2011. Late-replicating domains have higher divergence and diversity in Drosophila melanogaster. Mol Biol Evol 29: 873-882.

Wojtowicz W.M., Wu W. Andre I., Qian B., Baker D., Zipursky S.L. 2007. A vast repertoire of Dscam binding specificities arises from modular interactions of variable Ig domains. Cell 130: 1134-1145.

Yamagata M., Sanes J.R., Weiner J.A. 2003. Synaptic adhesion molecules. Curr. Opin. Cell Biol. 15: 621-632.

Zhang P. and Spradling, A.C. 1995. The Drosophila salivary gland chromocenter contains highly polytenized subdomains of mitotic heterochromatin. Genetics 139: 659-670.

Zhimulev, I.F., Semeshin, V.F., Kulichkov, V.A., Belyaeva, E.S. 1982. Intercalary heterochromatin in Drosophila. I. Localization and general characteristics. Chromosoma 87: 197-228

Zhimulev I.F. 1998. Polytene chromosomes, heterochromatin, and position effect variegation. Adv. Genet. 37: 1-566.

Zhimulev I.F, Belyaeva E.S, Makunin I.V, Pirrotta V, Volkova E.I, Alekseyenko A.A., Andreyeva E.N., Makarevich G.F., Boldyreva L.V, Nanayev R.A., Demakova O.V. 2003a. Influence of the SuUR gene on intercalary heterochromatin in Drosophila melanogaster polytene chromosomes. Chromosoma 111:377-398.

This Page Is Intentionally Left Blank

Chapter 3: Additional Insights Into Polyploidy, Polytene Chromosomes, Differential Replication, and Genomic Rearrangements

Author Contribution

This chapter is solely my own work. It includes additional data analysis and experiments that have not yet been published as well as further discussion on how my thesis works relates to the current status of the field.

Genomic Rearrangements From Differential Replication

Under replication is caused by stalled replication forks which do not finish DNA replication. SuUR is proposed to slow the progress of replication forks preventing complete DNA replication in regions of under replication (Sher et al. 2012; Kolesnikova et al. 2013). Similar quantitative analysis of whole ovary DNA from various Dp 1887 derivatives (Zhang and Spradling 1994) (Figure 7). These derivatives have rearrangements of the Dp1187 minichromosome, which include terminal deletions and internal deletions. The rearrangement of the genomic sequence results in differences in the copy number variation between the different versions of the mini chromosome. It is likely that the differences occur due to changes in the chromatin of the region affected resulting from a rearrangement of the sequence and chromatin in and flanking these regions. These stalled forks probably stimulate binding of variant histone 2a which leads to the repair of the free ends (Andreyeva et al 2008). SuUR mutants clearly have differences in ploidy and variant histone 2a signal (Figure 8). It would be interesting to see the location of the variant histone 2a with ChIP-Seq and see if it coincides with under replication as it does under the microscope and decreases in SuUR mutants. It

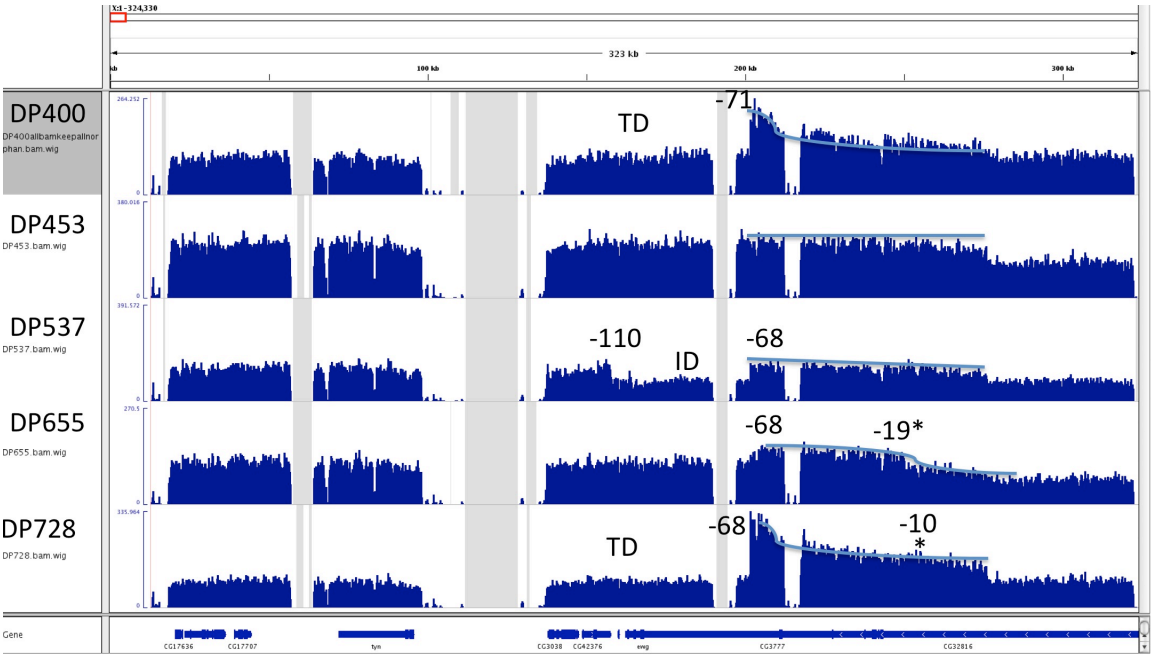


Figure 7. Raw Sequence Read Count of Dp 1887 Derivatives. Whole ovary DNA was prepared and sequenced as in chapter 2. Numbers indicate key features relative to the sc breakpoint (sc), such as deletions or a P-Element. These include terminal deletions (TD), internal deletions (ID), and no change. The blue curved line highlights the differences in copy number at different positions.

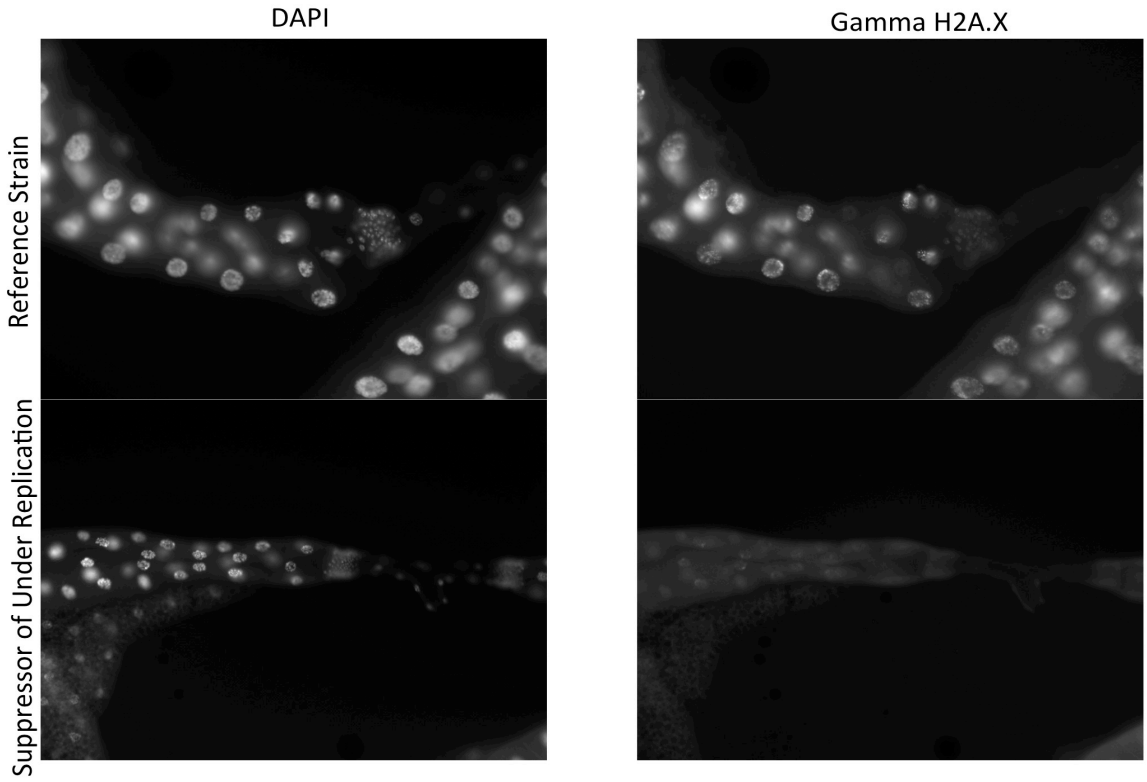


Figure 8. Gamma H2A.X staining of Polytene Chromosomes.

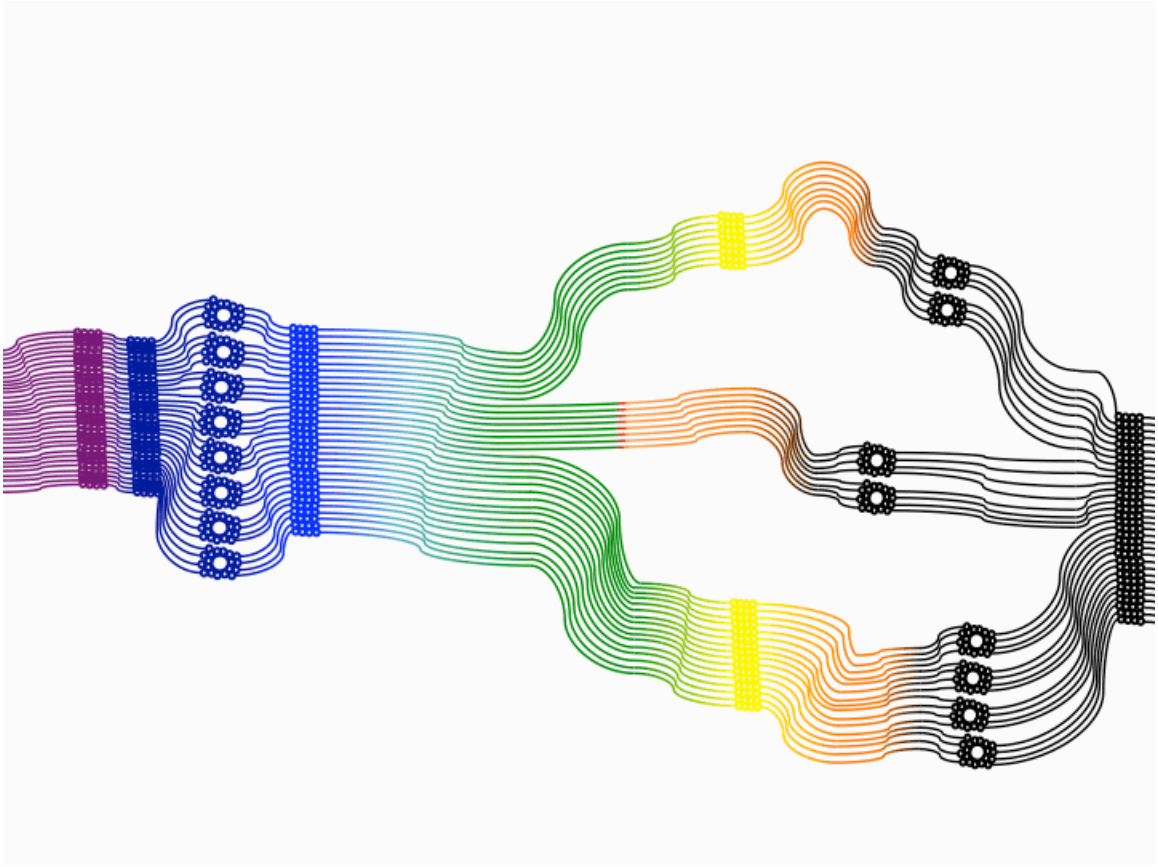
The reference strain and the suppressor of under replication mutant, as discussed in chapter 2, were stained with dapi and gamma H2A.X. Suppressor of under replication mutants were of lower ploidy and did not have intense gamma H2A.x staining.

Text continued from page 125

would be interesting to verify this in salivary gland DNA and determine of structural rearrangements are responsible for the variegation previously observed (Karpen et al. 1990).

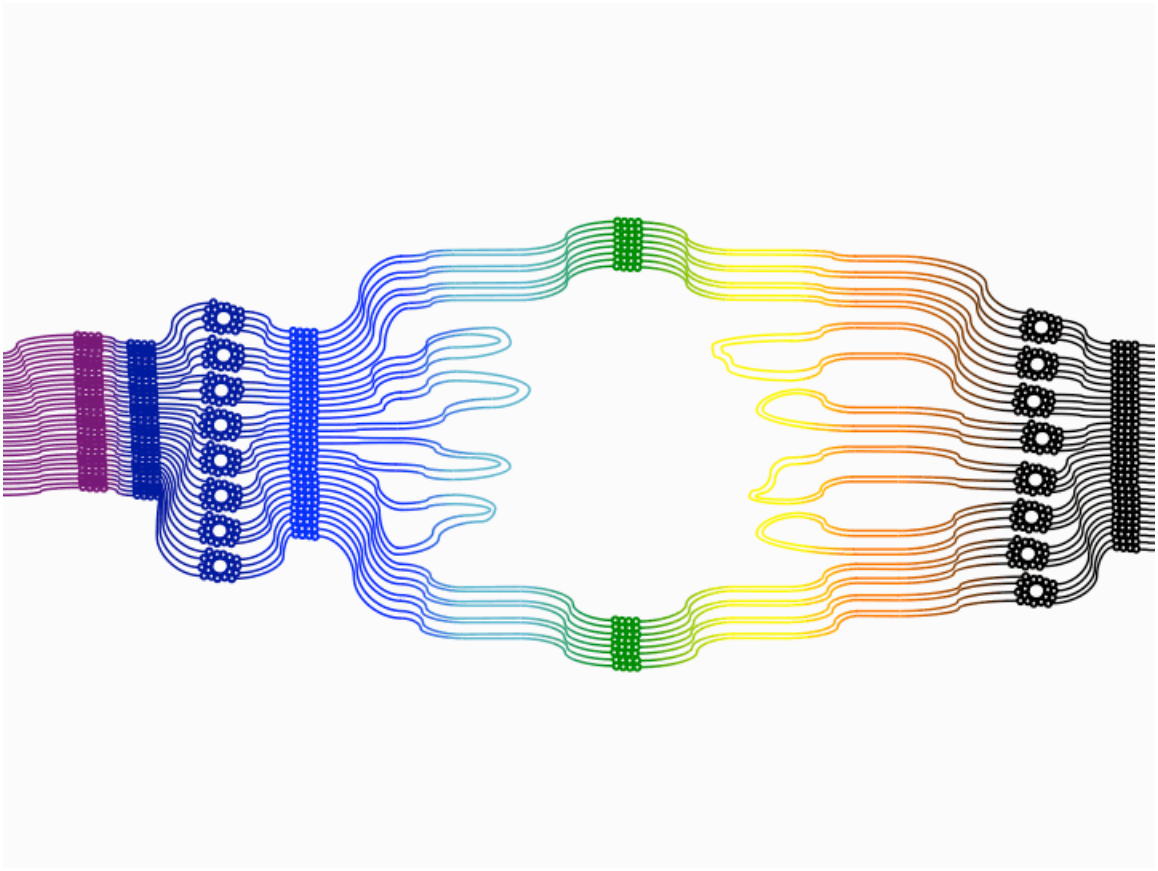
Structure of Under Replication Region

Interpreting hundreds of chromatin squashes and electron microscope images of 3rd instar larval salivary glands, and the frequency and nature of under replication and subsequent rearrangement as discussed in chapter 2, 3 images were drawn illustrating a single rearrangement between strands of the same and/or opposite orientation (Figure 9). Changes in the chromatin and therefore the bands occur. The ploidy of these images is lower than a fully developed 3rd instar larval salivary gland cell due to size limitations. These images provide insight into the wide range of structural changes that could occur to polytene chromosomes in regions of under replication.



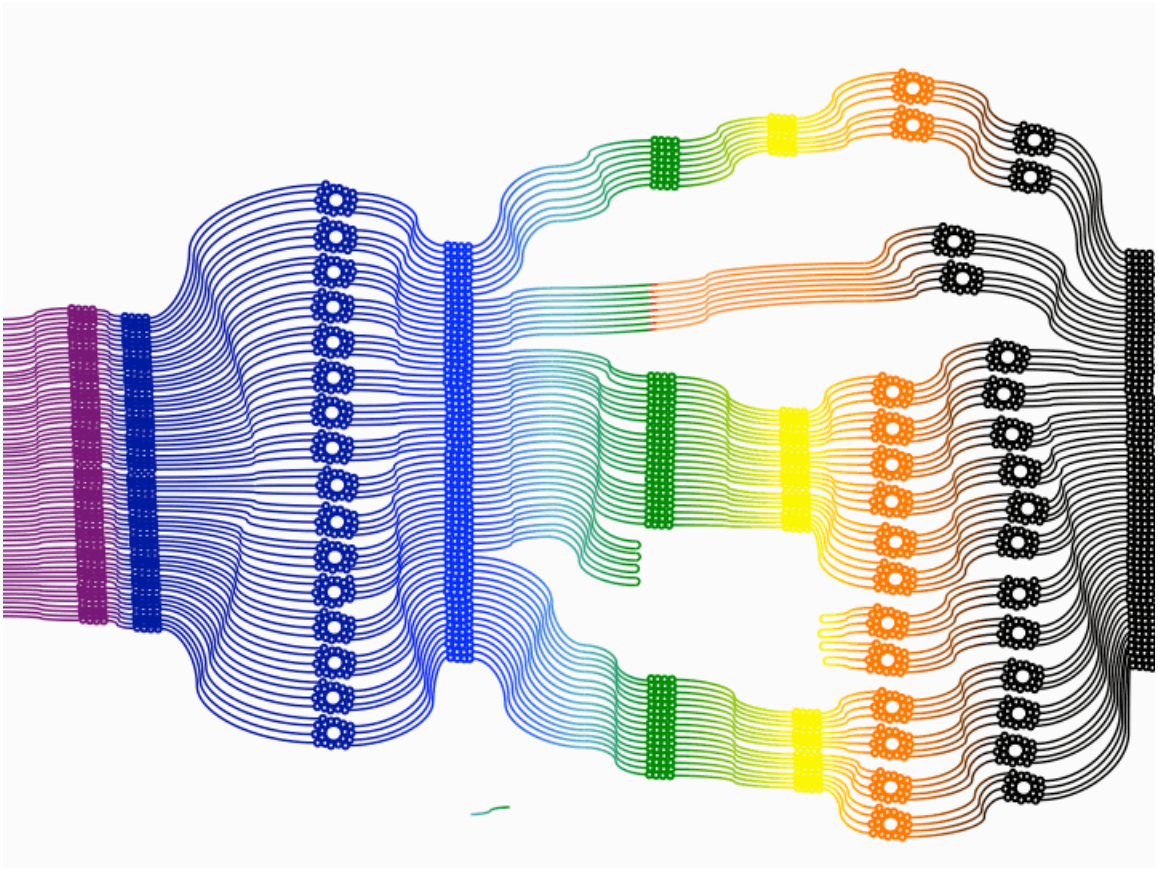
Rearrangement between strands of same orientation

Figure 9 continued from previous page



Rearrangement between strands of opposite orientation

Figure 9 continued from previous page



Rearrangement between strands of same and opposite orientation

Figure 9. Illustration of under replicated rearranged polytene chromosomes. Different colors represent different chromosome positions. The shapes represent different chromatin states along the genome. Each strand represents a double strand of genomic DNA.

References

Andreyeva E.N., Kolesnikova T.D., Belyaeva E.S., Glaser R.L., Zhimulev I.F. 2008. Local DNA underreplication correlates with accumulation of phosphorylated H2Av in the Drosophila melanogaster polytene chromosomes. Chromosome Res. 16: 851-62.

Karpen G.H. and Spradling A.C. (1990): Reducing DNA polydenization of a minichromosome region undergoing position-effect variegation in Drosophila. Cell. 63: 97-107.

Kolesnikova TD, Posukh OV, Andreyeva EN, Bebyakina DS, Ivankin A.V. and Zhimulev I.F. (2013) Drosophila SUUR protein associates with PCNA and binds chromatin in a cell cycle-dependent manner. Chromosoma 122: 55—66.

Sher N., Bell G.W., Li S., Nordman J., Eng T., Eaton M.L., Macalpine D.M., Orr-Weaver T.L. 2012. Developmental control of gene copy number by repression of replication initiation and fork progression. Genome Res 22: 64—75

Zhang P and Spradling A.C. (1994): Insertional mutagenesis of Drosophila heterochromatin with single P-elements. P.N.A.S. 91: 3539-3543.

This Page Is Intentionally Left Blank

Curriculum Vitae

William Henry Yarosh

Place and Date of Birth:

Houston, Texas. February 9, 1983.

Address:

Carnegie Institution of Washington

3520 San Martin Drive, Baltimore MD 21218

Education:

Irvine High School. Diploma. 1998-2001. Irvine, California.

University of California, Los Angeles. B.S. Biology. B.S. Microbiology, Immunology, and Molecular Genetics. 2001-2005. Los Angeles, California.

University of California, Irvine. M.S. Biology. 2005-2008. Irvine, California.

Johns Hopkins University. Ph.D. Cell, Molecular and Developmental Biology. 2009-2014. Baltimore, Maryland.

Work Experience:

University of California, Irvine. Teaching Assistant. 2006-2007. Irvine, California.

American Type Culture Collection. Virologist. 2008-2009. Manassas, Virginia.

Johns Hopkins University. Teaching Assistant. 2010-2014. Baltimore, Maryland.